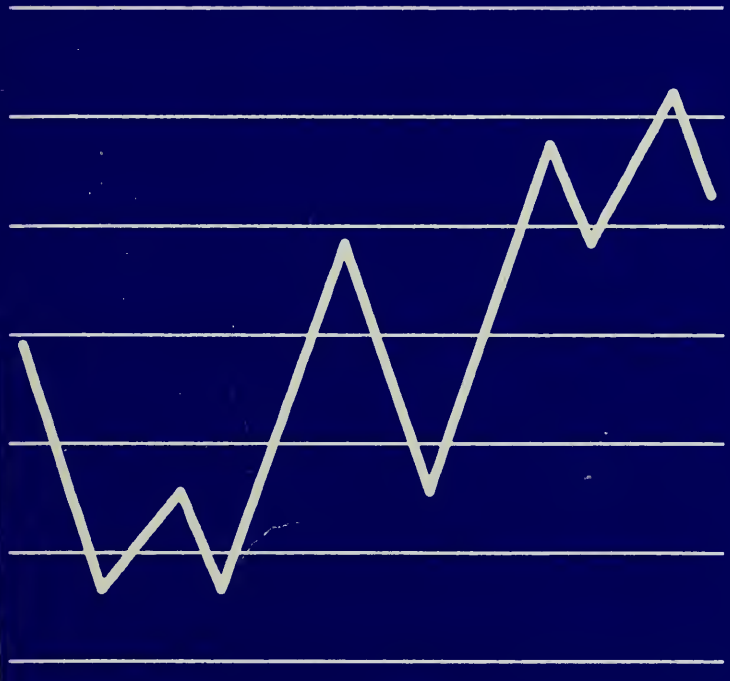
Agriculture Canada

MÉTHODES STATISTIQUES POUR LA GESTION DE LA QUALITÉ DES ALIMENTS

Agriculture Canada
MAY MAI - 7 1990
c 3
Library / Bibliothèque, Ottawa K1A 0C5



630.4
C212
P 5268
1990
fr.
c. 3

MÉTHODES STATISTIQUES POUR LA GESTION DE LA QUALITÉ DES ALIMENTS

Subhash C. Puri
Statisticien en chef
Direction générale de la production
et de l'inspection des aliments
Agriculture Canada, Ottawa

Agriculture Canada Publication 5268/F
On peut en obtenir des exemplaires à la
Direction générale des communications
Agriculture Canada, Ottawa K1A 0C7.

©Ministre des Approvisionnements et Services Canada 1990
N° de cat. A73-5268/1989F ISBN: 0-662-95662-1
Impression 1990 1M-4:90

Also available in English under the title:
Statistical methods for food quality management.

TABLE DES MATIÈRES

PRÉFACE

CHAPITRE 1 GESTION DE LA QUALITÉ GLOBALE 9

1.1 Introduction 9
1.2 Définitions et terminologie 10
1.3 Élaboration d'un programme de gestion de la qualité globale (GQG) 11
1.4 Éléments d'un programme de GQG 12
1.5 Programme de GQG pour un organisme de réglementation 14
1.6 Exemple: Programme d'assurance de la qualité pour les laboratoires d'analyses 15
1.7 Amélioration continue de la qualité 17
1.8 Mesure et amélioration de la productivité 18
1.9 Méthodes diagnostiques: Diagrammes cause-effet 19
1.10 Méthodes diagnostiques: Analyse de Pareto 20

CHAPITRE 2 MÉTHODES STATISTIQUES DE BASE 27

2.1 Introduction 27
2.2 Définitions 28
2.3 Organisation des données 28
2.4 Mesures statistiques 30
2.5 Probabilité 33
2.6 Permutations et combinaisons 33
2.7 Distributions statistiques 34
2.8 Distribution de la moyenne de l'échantillon et de la variance de l'échantillon 38
2.9 Distribution de t 39
2.10 Distribution de chi carré 39
2.11 Estimation: Intervalles de confiance 40
2.12 Vérification d'hypothèse 43
2.13 Régression linéaire et corrélation 48

CHAPITRE 3 MÉTHODES D'ÉCHANTILLONNAGE 53

3.1 Introduction 53
3.2 Types de procédés d'échantillonnage 53
3.3 Détermination de la taille d'un échantillon: Une formule 56
3.4 Méthodes de sélection d'un échantillon 57
3.5 Échantillonnage d'ensemble 60
3.6 Plan général d'une méthode d'échantillonnage 62

CHAPITRE 4 ÉCHANTILLONNAGE DE RÉCEPTION 65

4.1 Introduction 65
4.2 Classification des plans d'échantillonnage 65
4.3 Caractérisation d'un plan d'échantillonnage 66
4.4 Choix d'un plan d'échantillonnage 67
4.5 Systèmes de plans d'échantillonnage tout faits 69
4.6 Plans d'échantillonnage par attributs d'après le NQA: ISO/2859-1 69
4.7 Autres procédures d'échantillonnage de réception 75

CHAPITRE 5 CONTRÔLE STATISTIQUE DU PROCESSUS 79

5.1 Introduction 79
5.2 Types de graphiques de contrôle 80
5.3 Système de graphiques de Shewhart 80
5.4 Graphiques de contrôle de Shewhart: Formules et facteurs 81
5.5 Construction des graphiques de contrôle de Shewhart 83
5.6 Maîtrise et aptitude du processus 85
5.7 Graphiques de contrôle à somme cumulée 93

BIBLIOGRAPHIE 101

APPENDICE 105

Tableau 1	Aires sous la courbe normale des probabilités	106
Tableau 2	Points de pourcentage de la distribution de t	107
Tableau 3	Points de pourcentage de la distribution de chi carré	108
Tableau 4	Nombres aléatoires	109
Tableau 5A	ISO/2859-1: Lettres codes des échantillons	110
Tableau 5B	ISO/2859-1: Plans d'échantillonnage simple pour le contrôle normal	111
Tableau 5C	ISO/2859-1: Plans d'échantillonnage double pour le contrôle normal	112

PRÉFACE

"La recherche de l'excellence se poursuit . . ."

Un des aspects des plus fondamentaux de la gestion du secteur agro-alimentaire concerne la qualité et la sûreté des produits alimentaires et agricoles. Le producteur, le consommateur et l'organisme de réglementation sont également et collectivement intéressés par cet aspect et chacun vise à trouver des méthodes qui permettront d'assurer une grande qualité des produits alimentaires.

Pour atteindre l'excellence dans la production des aliments, il faut établir un système efficace de gestion de la qualité globale qui définit systématiquement les méthodes les plus précises possible. De plus, pour bien fonctionner dans le cadre de n'importe quel système de gestion de la qualité, quels que soient la nature, le type ou l'importance de celui-ci, le processus de prise de décision doit également pouvoir s'appuyer sur des compétences dans le domaine de l'analyse et de l'interprétation scientifiques des données générées par le système. Cet ouvrage a pour but de donner certaines lignes directrices qui permettront d'instaurer des méthodes d'assurance de la qualité ainsi que d'établir des outils statistiques de base pour l'analyse et l'interprétation des données. Les concepts et les techniques sont présentés très simplement et devraient aider toute personne qui se consacre à la gestion de la qualité et de la sûreté des produits alimentaires.

Je tiens à remercier Donna McGovern pour l'aide considérable qu'elle m'a apportée dans la rédaction de cet ouvrage, Roger Trudel qui m'a fait part d'excellentes observations professionnelles et Elise Benoit qui a participé à la conception de la page de titre. J'aimerais également exprimer ma profonde gratitude à ma femme Shashi et à mes deux filles Pamela et Anu pour leur patience et leur endurance.

Ottawa, Canada — Subhash C. Puri

À mes beaux-parents: Ram Parkash et Bimla Sahney

Digitized by the Internet Archive
in 2012 with funding from
Agriculture and Agri-Food Canada – Agriculture et Agroalimentaire Canada

http://www.archive.org/details/mthodesstatistiq00puri

CHAPITRE 1

GESTION DE LA QUALITÉ GLOBALE

1.1 Introduction

Les aliments sont essentiels à la survie de l'homme. Leur qualité représente donc un élément d'une extrême importance pour nous tous. Nous devons tous veiller à ce que nos produits agro-alimentaires soient sains, nutritifs, sûrs et sans danger pour notre santé. Nous pouvons atteindre cet objectif en fixant des normes de qualité et de sûreté élevées et en établissant des systèmes efficaces de gestion de la qualité globale pour voir au respect de ces normes.

Il y a généralement trois intéressés dans le secteur agro-alimentaire: le producteur, le consommateur et l'organisme de réglementation. Le consommateur définit les exigences en matière de qualité/sûreté, le producteur cherche à les satisfaire, et l'organisme de réglementation confirme que ces exigences ont été satisfaites. Chacun a un rôle important à jouer, mais l'excellence en matière de qualité des aliments ne peut être atteinte que grâce à la détermination, à l'engagement et à la coopération de tous ces intéressés.

Voici quels sont pour eux les points essentiels à considérer au chapitre de la qualité et de la sûreté des aliments:

- aliments propices à la consommation
- aliments sains, nutritifs et sûrs
- conformité aux spécifications, aux tolérances, aux normes et aux règlements établis en matière de quantité, de qualité et de contenu
- utilisation et divulgation correctes des pesticides, insecticides, herbicides, engrais, résidus, additifs, agents de conservation et autres produits chimiques

La production d'aliments de grande qualité est une responsabilité qui doit être partagée également par les trois intéressés. Ainsi, la responsabilité du producteur est tripartite:

- obtenir et maintenir la qualité réelle du produit de manière à satisfaire continuellement les exigences prévues du consommateur,
- fournir à ses propres directeurs l'attestation/l'assurance de la qualité recherchée,
- fournir à l'organisme de réglementation et à l'acheteur une attestation/assurance quant à la conformité du produit livré aux normes de qualité désirées. Selon les stipulations contractuelles, cela peut supposer qu'il faille en faire la démonstration aux intéressés.

L'organisme de réglementation a un double rôle à jouer: aider les producteurs à fournir des produits de grande qualité et protéger les consommateurs en vérifiant si les normes de qualité sont respectées. À cette fin, il devrait collaborer avec

les producteurs à l'élaboration et à l'instauration d'un système interne de gestion de la qualité globale, ainsi qu'à l'établissement d'un protocole d'assurance de la qualité propre à chaque producteur auquel celui-ci serait, en principe, tenu de se conformer.

Le consommateur, tout en n'étant que le destinataire du produit, doit faire en sorte de donner au producteur et à l'organisme de réglementation des informations en retour et toutes sortes de renseignements utiles sur la qualité, la quantité, le prix et la sûreté des produits alimentaires ainsi que sur les dangers que ceux-ci peuvent présenter.

Un système efficace de gestion de la qualité globale permet d'assigner les responsabilités et de trouver éventuellement la cause véritable d'un problème. Il constitue un mécanisme systématique d'analyse et d'amélioration continues de la qualité. Voici certains des avantages qui résultent de la mise en oeuvre appropriée d'un programme de GQG: constance et uniformité dans l'application des procédures; inspection du produit final moins longue; vérification; qualité et crédibilité accrues; réduction des coûts liés à l'évaluation, à la détection, à la prévention, aux défaillances internes, aux défaillances externes, aux pertes, à la non-conformité et à l'inspection; amélioration continue de la qualité et de la sûreté des produits; etc.

1.2 Définitions et terminologie

La gestion de la qualité globale (GQG) désigne la totalité des fonctions de gestion de la qualité et de la sûreté des aliments. À titre de stratégie corporative pour l'amélioration continue de la qualité, elle permet d'aborder de façon structurée et ordonnée l'identification et la solution des problèmes et propose un plan adaptable pour mettre en oeuvre les améliorations qui s'imposent. Les systèmes qualité portent différents noms: GQG (gestion de la qualité globale), MQG (maîtrise de la qualité globale), MQEE (maîtrise de la qualité à l'échelle de l'entreprise, utilisé par les Japonais), etc. Un système de GQG comprend fondamentalement trois composants:

- Gestion de la qualité (G.Q.): fonctions de gestion
- Maîtrise de la qualité (M.Q.): techniques et activités opérationnelles
- Assurance de la qualité (A.Q.): actions préétablies pour fournir la confiance appropriée

Voici une liste de définitions de certains termes relatifs à la qualité, que l'on peut trouver dans la Norme internationale, ISO/8402: "Qualité-Vocabulaire":

Qualité: Ensemble des propriétés et caractéristiques d'un produit ou d'un service qui lui confèrent son aptitude à satisfaire des besoins exprimés ou implicites.

Remarque: La qualité se définit également comme "l'adaptabilité à l'utilisation", "l'adaptabilité au but", "la conformité aux exigences", ou "la satisfaction du client". L'Association des normes japonaise la définit comme "la perte

transmise à la société à partir du moment où un produit est livré". La qualité est un état plutôt qu'une caractéristique spécifique d'une entité. Puisqu'elle est la somme de toutes les caractéristiques d'un produit, elle devrait les englober toutes.

Gestion de la qualité: Cet aspect de la fonction de gestion globale qui détermine et met en oeuvre la politique en matière de qualité.

Assurance de la qualité: Ensemble des actions préétablies et systématiques nécessaires pour donner la confiance appropriée en ce qu'un produit ou un service satisfera aux exigences données relatives à la qualité.

Maîtrise de la qualité: Techniques et activités à caractère opérationnel utilisées en vue de répondre aux exigences relatives à la qualité.

Système qualité: Ensemble de la structure organisationnelle, des responsabilités, des procédures, des procédés et des ressources pour mettre en oeuvre la gestion de la qualité.

1.3 Élaboration d'un programme de GQG

La responsabilité de fabriquer un bon produit dès le départ repose principalement sur le producteur. Celui-ci peut obtenir une meilleure qualité en établissant des contrôles efficaces à l'aide d'un programme de GQG. Grâce à un programme de GQG correctement mis en oeuvre et utilisé dans l'ensemble de son établissement, le producteur n'aura plus à consacrer autant de temps à l'inspection interne du produit final, et l'organisme de réglementation qui doit faire la vérification externe pourra aussi voir sa tâche allégée. Un système de gestion de la qualité globale est un véhicule qui permet d'améliorer continuellement la qualité.

On ne peut donner que les éléments généraux de la marche à suivre pour élaborer un programme de GQG. Ces éléments peuvent servir à mettre au point un programme de GQG unique adapté aux exigences d'une entreprise particulière. Par exemple, un programme de GQG d'une industrie manufacturière doit être constitué des éléments suivants:

- **Système de gestion de la qualité:** comprend tous les aspects de la gestion de la qualité comme la responsabilité de la direction, les systèmes de gestion, de maîtrise, d'établissement des coûts, d'évaluation et d'amélioration de la qualité, l'analyse du marché, la politique, la répartition des ressources, etc.
- **Maîtrise de la qualité:** comprend tous les points de maîtrise cruciaux ainsi que les aspects opérationnels et techniques de la maîtrise de la qualité pendant la production.
- **Assurance de la qualité de l'approvisionnement:** procédures d'assurance de la qualité visant à veiller à ce que les fournitures acquises soient de grande qualité.

- **Assurance de la qualité interne:** procédures d'évaluation internes visant à veiller à ce que le produit final soit de grande qualité.
- **Assurance de la qualité externe:** protocole contractuel d'assurance de la qualité qui garantit à l'organisme de réglementation et à l'acheteur, ou, en bout de ligne, au consommateur, que le produit livré est de grande qualité.

Voici les étapes de base à suivre pour élaborer un programme de GQG:

- Identifier la situation/le produit/l'entité qui nécessite l'établissement d'un programme de GQG.
- Préciser les buts et les objectifs.
- Dresser une liste complète des activités relatives à la situation.
- Classer les activités par catégories, soit gestion, systèmes, maîtrise, procédures, analyse, évaluation, vérification, mesures à prendre, informations en retour, amélioration, etc.
- Énumérer, pour chaque catégorie, toutes les mesures à prendre.
- Exposer les instructions détaillées à suivre pour effectuer chaque mesure comprise dans une catégorie.
- Le document final servira de protocole de GQG.

1.4 Éléments d'un programme de GQG

Vous trouverez ci-après une liste type qui englobe généralement tous les éléments d'un programme de GQG selon le point de vue d'un producteur. Il s'agit d'une liste générique que l'on peut adapter et utiliser pour concevoir un programme de GQG propre à une entité spécifique. Elle s'inspire des lignes directrices établies dans (i) la Norme nationale du Canada, CAN3-Z299: Programme d'assurance de la qualité — Association canadienne de normalisation, et dans (ii) la Norme internationale, ISO/9004: Gestion de la qualité et éléments de système qualité — Lignes directrices.

Éléments d'un programme de GQG — Liste type

- **Responsabilités de la direction:** politique, objectifs, planification, système de gestion, structure et responsabilités organisationnelles
- **Structure du système qualité:** responsabilités et autorité en matière de qualité, structure organisationnelle, ressources et personnel, procédures opérationnelles
- **Documentation du système qualité:** procédures et politiques relatives à la qualité, manuel qualité, plans qualité, dossiers qualité
- **Vérification du système qualité:** plan de vérification, vérification proprement dite, compte rendu des résultats de la vérification et suivi

- **Examen et évaluation du système de gestion de la qualité**
- **Considérations sur les coûts relatifs à la qualité:** choix des éléments appropriés; collecte et analyse des données sur les coûts; catégories de coûts: détection, évaluation, prévention, défaillance interne, défaillance externe; compte rendu sur les coûts fait à la direction
- **Qualité en commercialisation:** analyse du marché, descriptif relatif au produit, informations en retour du client
- **Qualité en définition et en conception:** préparation et objectif de la conception, essai et évaluation du produit, qualification et validation de la conception, analyse de la conception, conception de référence et examen de la production, examen d'aptitude à la mise en marche, maîtrise des modifications de conception, requalification de la conception
- **Qualité en approvisionnement:** exigences relatives aux spécifications et aux bons de commande, sélection des fournisseurs qualifiés, entente quant à l'assurance de la qualité, entente quant aux méthodes de vérification, dispositions relatives au règlement des différends liés à la qualité, préparation et maîtrise des contrôles à la réception, dossiers qualité relatifs à la réception
- **Qualité en production:** préparation en vue de la maîtrise de la production; aptitude du processus; fournitures, installations et milieux de travail
- **Maîtrise de la production:** maîtrise des matériaux et possibilité de retrouver ceux-ci, maîtrise et entretien du matériel, processus spéciaux, documentation, maîtrise des changements de processus, maîtrise de l'état des vérifications, maîtrise des matériaux non-conformes
- **Vérification des produits:** matériaux et matériel à l'entrée, inspection en cours de fabrication, vérification du produit fini
- **Maîtrise du matériel de mesure et d'essai:** maîtrise des instruments de mesure, éléments de maîtrise, maîtrise des instruments de mesure chez les fournisseurs, mesures correctives, essais externes
- **Non-conformité:** identification, séparation, examen, destination, documentation, prévention des récurrences
- **Mesures correctives:** attribution des responsabilités, évaluation de l'importance/priorité, recherche des problèmes possibles, analyse du problème, mesures préventives, maîtrise du processus, destination des articles non-conformes, changements permanents
- **Manutention et activités faisant suite à la production:** manutention, entreposage, identification, emballage et livraison; service après vente; retour d'information du marché et surveillance du produit
- **Documentation relative à la qualité:** spécifications, instructions concernant l'inspection, procédures d'essai, instructions de travail, manuel qualité, procédures opérationnelles, procédures d'assurance de la qualité, etc.

- **Dossiers qualité:** rapports d'inspection, données d'essai, rapports de qualification, rapports de validation, rapports de vérification, rapports sur l'analyse des matériaux, données d'étalonnage, rapports sur les coûts relatifs à la qualité, etc.
- **Personnel:** formation, compétences, appréciation, motivation
- **Sûreté du produit et responsabilité du fait du produit:** normes de sûreté appropriées; déclaration quant à la qualité, à la quantité et au contenu; avertissement à l'intention de l'utilisateur concernant les risques liés au produit; possibilité de retrouver le produit comme garantie sécurité
- **Utilisation de méthodes statistiques:** analyse du marché, maîtrise du processus, niveau de conformité, qualité moyenne de production, analyse des données, évaluation de la sûreté et analyse des risques, procédures d'échantillonnage statistique, graphiques de contrôle de la qualité, méthodes de conception, évaluation des performances, établissement/modification des tolérances, etc.

1.5 Programme de GQG pour un organisme de réglementation

On peut procéder de la même façon pour élaborer un programme de GQG destiné à un organisme de réglementation qui oeuvre dans le domaine agro-alimentaire en sélectionnant les éléments appropriés sur la liste type ci-dessus. Une telle liste générique pourrait comprendre les éléments suivants:

- Responsabilités de la direction
- Structure du système qualité
- Documentation du système qualité
- Plans de surveillance, d'inspection et de vérification
- Description des exigences en matière de tolérances, de spécifications et de réglementation
- Documents relatifs à la qualité et dossiers qualité
- Mesures à prendre en cas de non-conformité
- Mesures correctives
- Utilisation de méthodes statistiques pour l'analyse des données

Pour chacun de ces éléments principaux, on dresse également une liste de sous-éléments à laquelle on joint des instructions détaillées portant sur chacun des sous-éléments. Voici certaines considérations opérationnelles essentielles à incorporer dans les sous-éléments:

- Déterminer le niveau de risque comparatif lié à chaque produit.
- Préciser, pour chaque niveau ou catégorie de risque, la fréquence appropriée du niveau d'inspection.

- Déterminer, pour chaque niveau de risque, volume de production et état de la qualité, la fréquence d'inspection du produit.
- Préciser, pour chaque inspection, la taille de l'échantillon en rapport avec le niveau de qualité acceptable et le taux de conformité établi.
- Pour chaque inspection par échantillonnage, s'assurer de l'homogénéité des lots et du caractère aléatoire/représentatif des procédures d'échantillonnage.

1.6 Exemple: Programme d'assurance de la qualité pour les laboratoires d'analyses

Il est relativement facile, à partir des lignes directrices et des éléments de programme décrits dans la Section 1.4, d'élaborer un protocole d'assurance de la qualité pour une entité quelconque. Prenons à titre d'exemple d'application l'établissement d'un programme d'assurance de la qualité pour un laboratoire d'analyses. On prendra tout d'abord en considération le plan d'action suivant:

- Établir un profil des exigences en matière d'assurance de la qualité globale.
- Évaluer les pratiques courantes du laboratoire relativement à ces exigences.
- Préciser la marche à suivre pour appliquer les exigences en matière d'assurance de la qualité au laboratoire.
- Enseigner les pratiques et la marche à suivre nouvelles ou modifiées au personnel du laboratoire et former le personnel en conséquence.
- Établir un système de gestion de l'évaluation périodique pour veiller à ce que le programme soit vraiment efficace.

L'étape suivante consiste à dresser une liste des éléments du programme d'assurance de la qualité qui engloberait toutes les activités du laboratoire. Conformément aux lignes directrices de la Section 1.4, la liste comprendrait les éléments essentiels suivants:

- Plan d'assurance de la qualité
- Énoncés de politique(s)
- Objectifs
- Planification de la qualité
- Coûts relatifs à la qualité
- Tenue des dossiers et maîtrise des documents
- Série de procédures relatives à la garde des documents
- Formation pertinente
- Approvisionnement et maîtrise
- Réactifs et normes de référence
- Calibrage des instruments

- Entretien préventif
- Échantillonnage — identification et maîtrise de l'échantillon
- Validation des données et des méthodes
- Analyse en laboratoire et maîtrise de celui-ci
- Essais entre laboratoires et à l'intérieur du laboratoire
- Procédures de maîtrise statistique de la qualité
- Mécanisme des mesures correctives
- Consignes de sûreté
- Plan et fonctionnement du laboratoire
- Vérification du rendement et des systèmes
- Comptes rendus à la direction

L'établissement d'un manuel traitant de l'assurance de la qualité constitue une autre partie essentielle du protocole d'assurance de la qualité. Ce manuel énonce clairement les méthodes et les consignes d'exploitation spécifiques à suivre pour répondre aux besoins du laboratoire et pour que celui-ci atteigne ses objectifs en matière de qualité. Un manuel qui traite d'assurance de la qualité est un recueil de documents sur lesquels le personnel de laboratoire peut s'appuyer pour faire un bon travail. Il indique la politique, l'organisation, les objectifs, les activités fonctionnelles et les activités ayant spécifiquement trait à l'assurance de la qualité qui ont été arrêtés pour permettre au laboratoire d'atteindre le niveau de qualité qu'il s'est fixé.

Voici une présentation typique d'un manuel qui traite d'assurance de la qualité et qu'on destine à un laboratoire:

- Page de titre, des espaces étant prévus pour les signatures autorisées
- Table des matières
- Organisation du laboratoire et ses responsabilités
- Structure organisationnelle
- Plan d'assurance de la qualité et objectifs
- Système d'assurance de la qualité
- Mesures correctives
- Formulaires
- Procédures d'évaluation de la qualité
- Comptes rendus de l'assurance de la qualité présentés à la direction
- Liste de diffusion

1.7 Amélioration continue de la qualité

La qualité n'est pas une question de tactique mais de stratégie. C'est l'affaire de tout le monde et la qualité ne peut pas être fabriquée. La qualité est infusée à un produit et scellée dans celui-ci d'une façon systématique. En deux mots, la qualité est un objectif de longue haleine, non une fonction de courte durée. Par conséquent, pour atteindre les objectifs fixés, il est essentiel d'établir d'abord un système de GQG efficace et d'élaborer par la suite un programme de surveillance pour l'amélioration de la qualité. Voici certains composants et éléments essentiels d'un programme d'amélioration de la qualité:

Le programme:

- Engagement, participation et soutien complets et actifs de la direction
- Participation complète, très visible et orientée vers l'action de tous les intéressés, et intention manifeste de ceux-ci de produire un travail de qualité
- Sensibilisation et engagement complets du travailleur
- Respect et reconnaissance du travailleur
- Maîtrise de la qualité en fonction du client
- Objectif stratégique de longue haleine

Le système:

- Équipes qui se consacrent à l'amélioration de la qualité
- Maîtrise de la qualité adaptée à chaque projet
- Orientation professionnelle et formation des directeurs

Les méthodes:

- Outils statistiques utilisés par les directeurs pour prendre des décisions: distribution de fréquence, histogramme, diagramme cause-effet, analyse de Pareto, analyse du coût économique, etc.
- Méthodes formelles, structurées et ordonnées: analyse de la tâche, analyse du problème, analyse de la cause fondamentale, mesures correctives/ préventives, etc.
- Maîtrise statistique du processus

On dressera aussi, en se fondant sur les questions suivantes, une liste de questions type qui est en rapport avec les éléments du programme décrit ci-dessus:

- Tous les intéressés ont-ils été informés clairement des buts et objectifs et les ont-ils bien compris?
- Les fonctions et responsabilités attribuées étaient-elles appropriées, et ont-elles été clairement définies, bien comprises, et acceptées?

- A-t-on distribué les ressources (humaines, financières et technologiques) de façon appropriée et optimale?
- Y a-t-il des écarts entre les résultats obtenus et les résultats prévus?
- Qui est responsable de ces écarts?
- Quelles mesures doit-on prendre pour obtenir les résultats prévus et pour chercher à améliorer encore plus la qualité?
- Y a-t-il lieu d'établir un mode de vérification automatique de l'état périodique du programme?

Afin d'évaluer l'efficacité du programme, nous recommandons de procéder comme suit:

- Désigner une équipe de projet; préciser le problème; établir les théories et les causes principales; élaborer des mesures correctives.
- Mettre en application et communiquer un plan d'action.
- Choisir un point contesté; déterminer les caractéristiques à évaluer; recueillir des données pertinentes.
- Analyser les données à l'aide de méthodes statistiques pertinentes.
- Vérifier les résultats; évaluer les améliorations; confirmer la suppression des causes principales comme prévu.
- Normaliser; établir/réviser les procédures; analyser d'autres problèmes.

1.8 Mesure et amélioration de la productivité

La productivité et la qualité sont des concepts inséparables. Nous devons mesurer la productivité comme notre aptitude à fournir des produits et des services de grande valeur qui répondent aux exigences du client, et ce à un coût minimum. Les trois facteurs prioritaires sont la sûreté, la qualité et le coût.

On peut définir la productivité en se fondant sur la qualité, l'opportunité et la rentabilité qui caractérisent la manière adoptée par un organisme pour accomplir sa mission. Elle est une mesure du degré d'efficacité avec laquelle les ressources sont coordonnées et utilisées dans le but d'obtenir un ensemble de résultats.

La productivité et la production sont deux choses différentes. Une meilleure production ne signifie pas nécessairement une meilleure productivité. Alors que la productivité se rapporte à l'utilisation efficace des ressources, la production correspond au mécanisme et (ou) aux méthodes de production de biens et de services. Une usine productive est celle qui a un volume de production élevé en termes de quantité de matériaux, d'énergie, de main-d'oeuvre, de capital et d'autres ressources utilisées.

La productivité est une combinaison d'efficacité et d'efficience que l'on exprime sous forme d'indice:

$$\text{Indice de productivité} = \frac{\text{rendement des ressources}}{\text{apport de ressources}} = \frac{\text{efficacité}}{\text{efficience}}$$

L'efficacité indique dans quelle mesure un objectif est atteint; on ne s'attache dans ce cas qu'à obtenir les résultats désirés sans trop se soucier des coûts. L'efficience définit dans quelle mesure les ressources disponibles sont bien utilisées en vue d'obtenir les résultats fixés, et tient compte du coût total de toutes les ressources nécessaires.

Craig et Harris (1973) recommandent un indice de productivité plus complexe comme suit:

$$P_t = O_t / (L_t + C_t + R_t + Q_t)$$

où

P_t = la mesure de la productivité pour la période t

O_t = la production totale de bons éléments pendant la période t (mesurée en dollars sous évalués ou de l'année de base)

L_t, C_t, R_t, Q_t = la valeur du dollar de l'année de base correspondant respectivement à la main-d'oeuvre, au capital, aux matériaux bruts ainsi qu'aux biens et services utilisés pendant la période t

Afin d'améliorer la productivité, il faut surveiller et analyser l'indice de productivité. Cette analyse est effectuée à l'aide de graphiques de contrôle qui seront décrits dans le Chapitre 5.

1.9 Méthodes diagnostiques: Diagrammes cause-effet

Un processus est un ensemble de conditions ou de causes qui sont réunies pour produire un résultat ou un effet donné. Les causes correspondent presque toujours aux six facteurs suivants: les matériaux, la machine, la personne, la méthode, l'argent et la direction. Le **diagramme cause-effet (C-E),** conçu par le professeur japonais Kaoru Ishikawa en 1950, constitue un excellent moyen d'étudier le rapport entre des causes et un effet. Ce diagramme porte également le nom de **diagramme de remue-ménage** ou **diagramme en arête de poisson.** Un diagramme C-E de base est illustré à la Figure 1.1. Ce type de diagramme sert d'outil diagnostique pour identifier le problème ou l'effet, déterminer toutes les causes, évaluer les procédures opérationnelles, trouver différentes façons de résoudre le problème, et améliorer la qualité du processus.

Construction d'un diagramme C-E

- Déterminer les caractéristiques de la qualité ou les effets dont il faut trouver les causes.

- Tracer une ligne horizontale qui se termine à droite par une pointe de flèche devant laquelle on dessine un carré pour indiquer l'effet ou le problème.
- Écrire les principaux facteurs ou causes, c.-à-d., les matériaux, la personne, la méthode, la machine, l'argent, et la direction, en joignant chacun de ces facteurs à la flèche horizontale à l'aide d'une flèche inclinée pointant vers celle-ci.
- Ajouter des ramifications-flèches pour identifier les sous-causes de la cause principale. Procéder de la même façon pour ajouter des sous-sous-causes.
- S'assurer que toutes les causes possibles ont été prises en compte.
- Le graphique ou le tableau final correspond à un diagramme C-E.

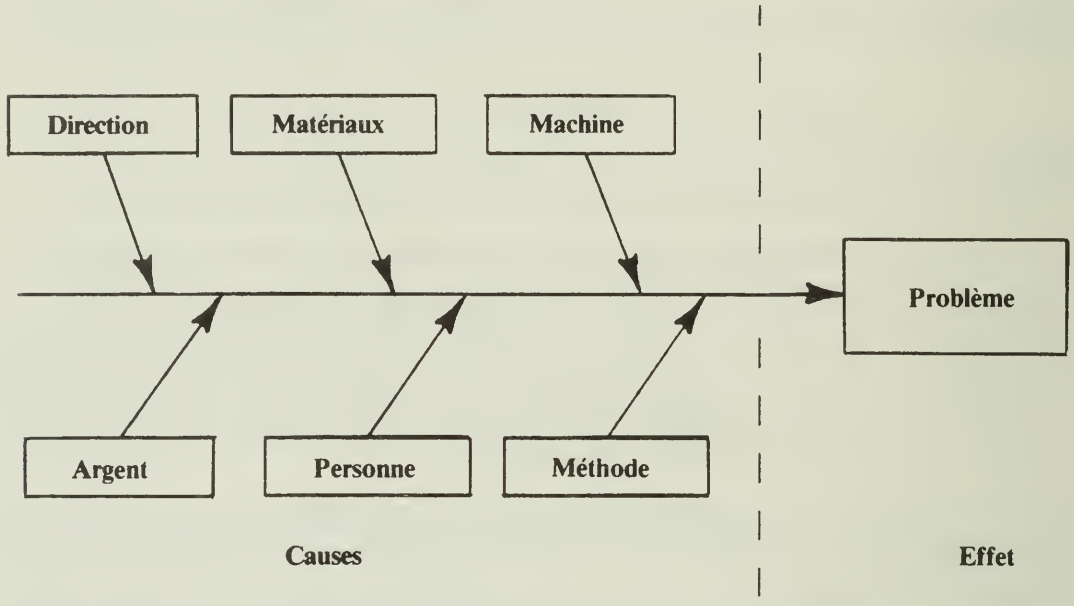


Figure 1.1 Diagramme cause-effet identifiant les principales causes

À titre d'exemple, vous trouverez à la Figure 1.2 une représentation détaillée d'un diagramme C-E illustrant un problème causé par des "défauts de serrage des capsules" sur la chaîne de traitement d'une usine de conditionnement des aliments. Après avoir identifié le défaut, on a fait une recherche détaillée de toutes les causes possibles qui ont été diagnostiquées à l'aide du diagramme C-E. Il était alors très facile de déterminer les causes et de les classer par ordre d'importance pour ainsi permettre à la direction de prendre une décision rapide et judicieuse quant aux mesures correctives/préventives qui s'imposaient.

1.10 Méthodes diagnostiques: Analyse de Pareto

L'analyse de Pareto constitue une autre méthode diagnostique utile qui s'ajoute à celle des diagrammes cause-effet et qui permet de classer les causes de non-conformité par ordre d'importance et de les analyser. Cette technique repose principalement sur la prémisse suivante: tous les événements ou toutes les causes ne

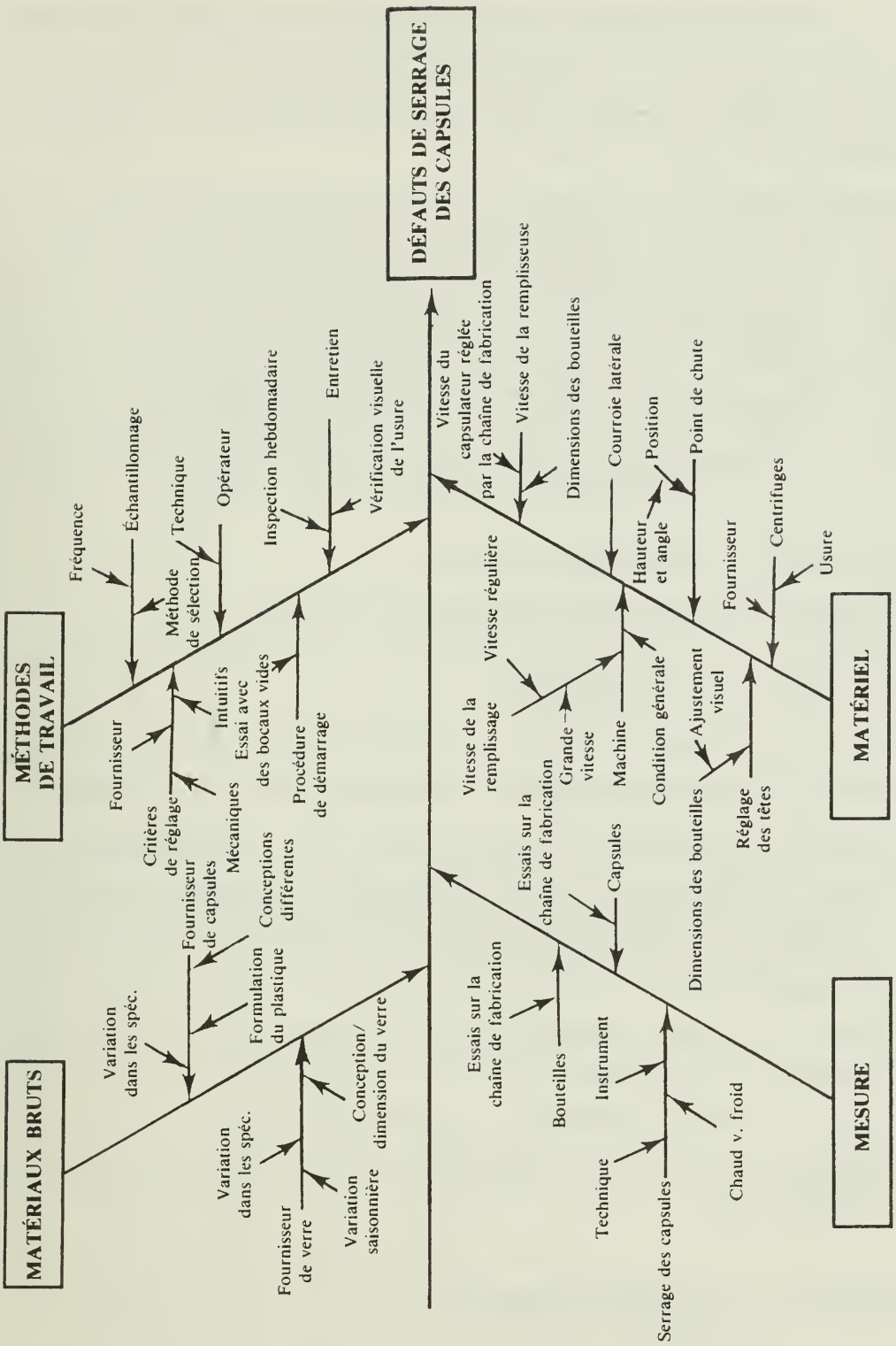


Figure 1.2 Diagramme cause-effet: Défauts de serrage des capsules

sont pas distribués uniformément pour autant que l'on considère leurs effets. Ils sont mal distribués en ce sens que relativement peu de causes sont à l'origine de la plupart des effets. Le principe de Pareto parle dans ce cas du phénomène 80-20, qui suppose qu'environ 80% des non-conformités sont généralement attribuables à seulement 20% des causes.

Le concept original de la distribution de Pareto a été élaboré par Vilfredo Pareto, un sociologue-économiste italien du dix-neuvième siècle, qui s'est penché sur la mauvaise répartition des richesses et des revenus. Il était d'avis que 80% des richesses d'un pays étaient normalement contrôlées par 20% de la population. Juran a repris cette idée en 1964 et l'a appliquée à la maîtrise de la qualité. Il a proposé de diviser les causes de non-conformité en deux catégories, les **'quelques causes fondamentales'** et le **'grand nombre de causes négligeables'**, puis de les classer par ordre d'importance de manière à pouvoir prendre des mesures pour corriger les **quelques causes fondamentales** qui sont à l'origine des principales pertes dûes à la non-conformité.

Construction d'un diagramme de Pareto

- Énumérer tous les éléments essentiels du processus.
- Établir le mode de classification des données, c.-à-d., type de défaut, numéro de pièce, équipe, opération, etc.
- Choisir une période appropriée et recueillir toutes les données pertinentes.
- Inscrire, pour chaque catégorie, la fréquence totale de l'occurrence.
- Ordonner les éléments en fonction de cette mesure, non de leur classification.
- Tracer un graphique à barres des fréquences, en commençant à gauche par la catégorie de la fréquence la plus élevée pour finir à droite par la catégorie présentant la fréquence la plus faible. Un bon graphique comporte cinq ou six catégories.
- Tracer sur le même graphique le graphe linéaire représentant la fréquence cumulée de chaque catégorie.
- Donner un titre au graphique et ajouter une légende.
- Prendre les mesures qui s'imposent pour corriger les 'quelques causes fondamentales'.

Exemple 1.1 Analyse de Pareto

Prenons l'exemple du traitement d'un produit "mayonnaise" dans une usine de conditionnement des aliments. L'analyse de fin d'exercice des chiffres des coûts nous révèle que d'importantes pertes sont imputables à des défauts de qualité. Le Tableau 1.1 indique, par ordre d'importance, les différentes catégories de coût; le diagramme de Pareto correspondant est illustré à la Figure 1.3.

TABLEAU 1.1: Pertes imputables à des défauts de qualité

Ordre d'importance	Catégorie	Coût ($)	Coût en pourcentage
1	Durée d'immobilisation de la chaîne	30,000	39.6
2	Pertes attribuables au contenant lui-même et à la façon dont il est fermé	18,000	23.7
3	Quantité renversée	15,000	19.7
4	Rectification des lots	5,000	6.5
5	Dommages attribuables à la manipulation des matériaux	4,000	5.3
6	Nouveau mélange	3,000	3.9
7	Correctifs apportés suite aux plaintes des clients	1,000	1.3
	Total	76,000	100.0

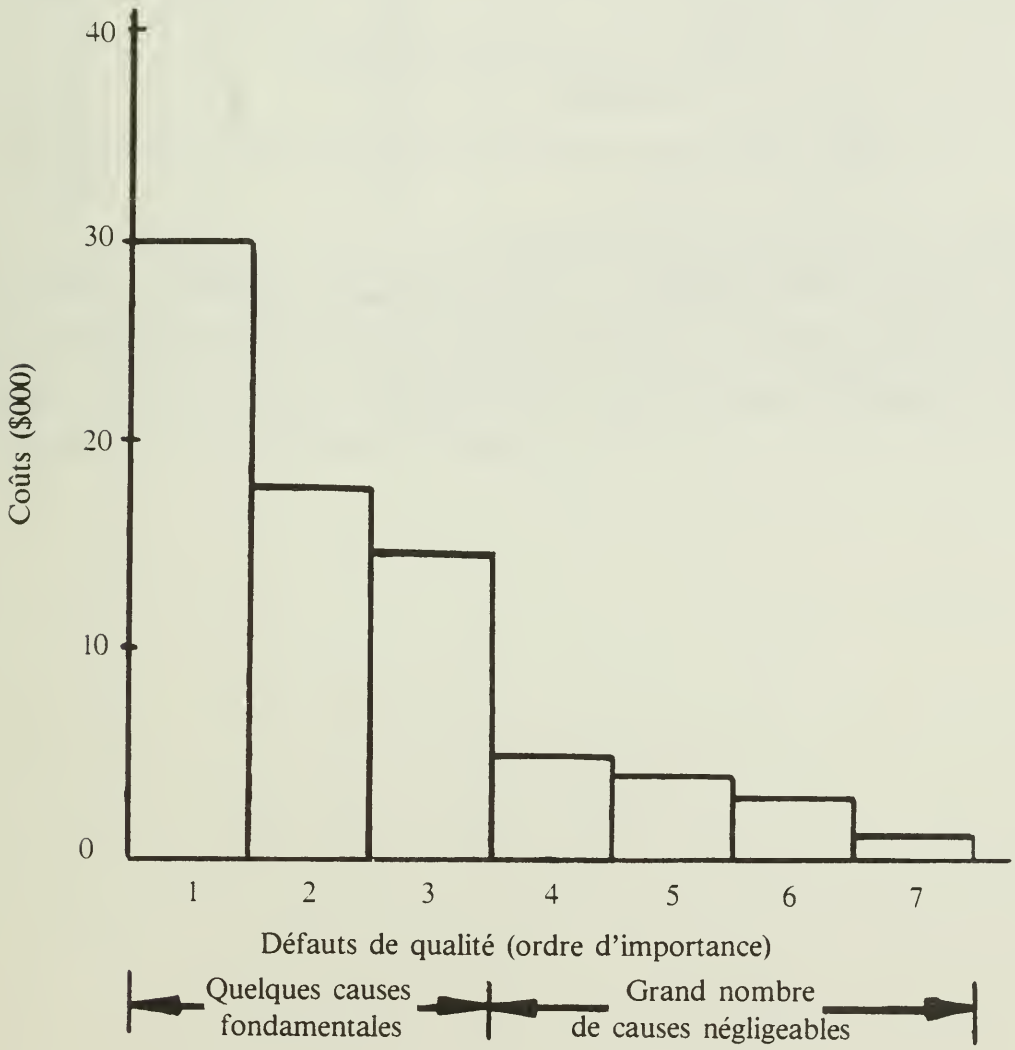


Figure 1.3 Diagramme de Pareto représentant les données du Tableau 1.1

Comme on peut le constater à partir de cette analyse, la durée d'immobilisation moyenne de la chaîne semble être la principale cause des pertes. On peut pousser plus loin l'analyse de cette catégorie en identifiant les causes et en les classant par ordre d'importance suivant l'exemple du Tableau 1.2 et du diagramme de Pareto correspondant de la Figure 1.4.

TABLEAU 1.2: Durée d'immobilisation moyenne (en minutes par équipe) de la chaîne d'emballage (pour une période de trois semaines)

Ordre d'importance	Catégorie	Durée d'immob. (min.)	Durée d'immob. (%)	Durée d'immob. cumulative (%)
1	Problème d'emballeuse	58	30.1	30.1
2	Problème d'étiqueteuse	47	24.4	54.5
3	Pas de verre	39	20.2	74.7
4	Pas de produit	12	6.2	80.9
5	Pas de capsule	10	5.2	86.1
6	Problème de capsulateur	9	4.7	90.8
7	Verre cassé	8	4.1	94.9
8	Pot de colle vide	6	3.1	98.0
9	Condition de la colle	3	1.5	99.5
10	Collage des caisses à l'aide de ruban adhésif	1	0.5	100.0
	Total	193	100.0	

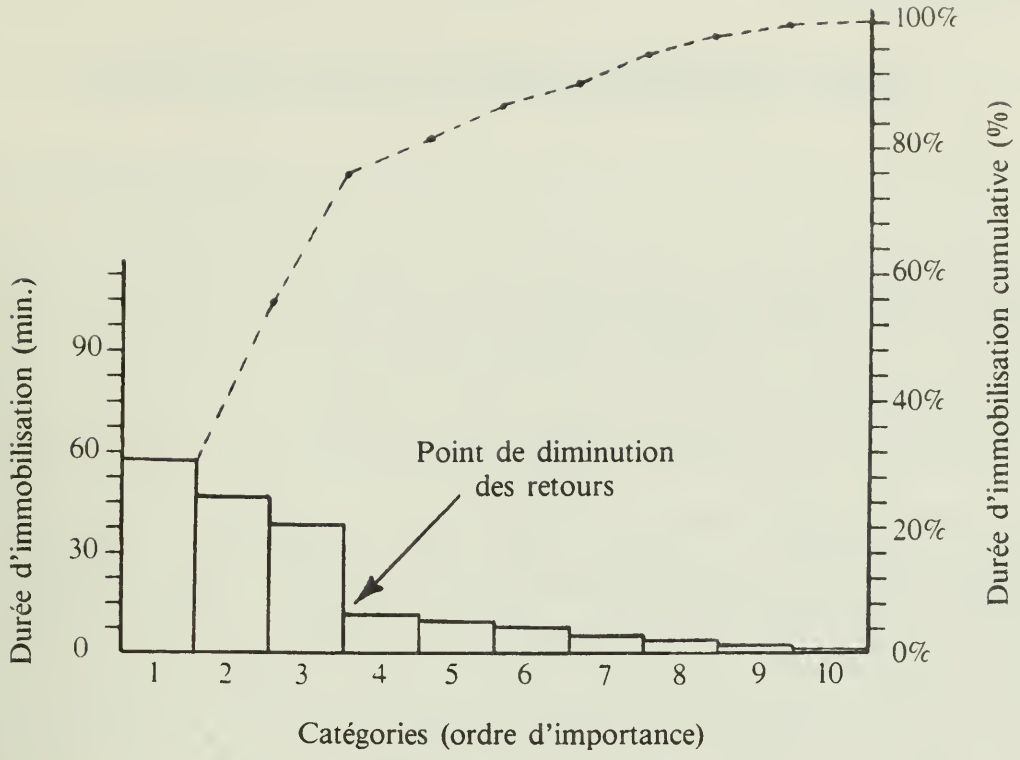


Figure 1.4 Diagramme de Pareto représentant les données du Tableau 1.2

En se fondant sur cette analyse, la direction peut facilement prendre une décision quant au problème auquel elle devrait s'attaquer en premier. Puisqu'il n'est pas rentable d'essayer de prévenir tous les problèmes, l'analyse de Pareto permet de bien définir l'ordre dans lequel les mesures correctives devraient être appliquées. Il reste qu'en matière de gestion de la qualité la véritable difficulté consiste à empêcher qu'un problème ne se pose.

CHAPITRE 2

MÉTHODES STATISTIQUES DE BASE

2.1 Introduction

Chaque enquête scientifique produit des informations et des données numériques. L'examen répétitif de chaque dossier de données brutes et inorganisées est généralement fastidieuse. Pour que la direction soit en mesure de prendre de bonnes décisions, la description et le résumé des résultats doivent être présentés de façon concise. À cette fin, les méthodes statistiques offrent des outils scientifiques qui permettent d'organiser, de résumer, d'analyser et d'interpréter les résultats.

D'un point de vue statistique, chaque valeur d'un groupe de mesures (échantillon) est assimilée à une réalisation unique d'une population hypothétique et infinie de mesures analogues. Même si, de façon générale, les membres de cet échantillon sont tous soumis à la mesure d'une même propriété d'un même lot ou population, on ne peut s'attendre qu'ils soient tous identiques. Les différences qu'on observe sont attribuables à des facteurs aléatoires et à une multitude d'autres facteurs identifiables associés aux techniques de mesure. L'objectif fondamental des statistiques consiste donc à déterminer les causes des variations, à évaluer leur importance et à proposer finalement des moyens de faire des inférences, à partir d'un échantillon, d'une population.

Voici certaines des questions qui sont le plus fréquemment posées au cours d'une étude, d'une enquête ou d'une expérience:

- Comment concevoir une expérience statistique valable?
- Quelle doit-être la taille de l'échantillon si l'on veut que celui-ci soit rentable, réalisable, valable statistiquement et suffisant pour obtenir des estimations fiables et un critère sur lequel se fonder pour prendre de bonnes décisions?
- Comment faire en sorte qu'un échantillon soit homogène et aléatoire, ainsi que représentatif du lot ou de la population?
- Comment estimer les paramètres d'une population à partir des statistiques d'un échantillon présentant un degré de confiance élevé?
- Comment analyser les différences entre plusieurs résultats d'échantillonnage?
- Comment effectuer une analyse de régression et de corrélation sur laquelle on peut s'appuyer pour faire des prévisions?
- Comment établir des méthodes de contrôle qui nous permettent de parvenir à la constance, à l'uniformité, à la répétabilité et à la reproductibilité des résultats?

2.2 Définitions

Population ou lot: l'ensemble du groupe d'unités qui fait l'objet de l'étude, et pour lequel des conclusions générales peuvent être dégagées.

Échantillon: une fraction de la population. Un échantillon devrait être représentatif de la population et être choisi au hasard. Un **échantillon aléatoire simple** est celui qui est obtenu par un procédé de sélection aléatoire de façon que chaque élément de la population ait une probabilité égale et indépendante de faire partie de l'échantillon.

Types de données:

- **Discrètes:** les données discrètes ne peuvent prendre que certaines valeurs (généralement entières) et supposent des comptages, par ex., le nombre de vaches d'une exploitation agricole, le compte d'éléments qui n'entrent pas dans un classement requis, etc. Lorsque des caractéristiques semblables sont dichotomiques (c.-à-d., défectueuses — non défectueuses), on parle **d'attributs.**
- **Continues:** données provenant d'une mesure ou d'une autre méthode d'estimation numérique; ces caractéristiques donnent des **variables,** par ex., la température, les valeurs pH, la production agricole, etc.

2.3 Organisation des données

Les données statistiques obtenues à la suite d'une étude scientifique consistent habituellement en un grand nombre d'observations. Pour donner une valeur informative à cette masse inorganisée de données, il faut les résumer, les décrire et les présenter. Le moyen de représentation visuel des données le plus répandu est **l'histogramme** ou **graphique à barres.** Appliqués aux données discrètes, ces graphiques se tracent généralement sans difficulté. En revanche, les données continues comme le poids, la température, la pression, la longueur et le pourcentage de matière sèche ne sont pas déjà groupées en catégories naturelles et doivent donc faire au préalable l'objet d'un certain classement. Pour y parvenir, on a recours à une **table des fréquences** dans laquelle l'étendue des données est répartie entre un nombre modéré de catégories et chaque observation est attribuée à une seule de ces catégories. Le nombre de catégories est arbitraire mais, en pratique, on s'arrange pour que $k = \sqrt{n}$, k représentant le nombre de catégories et n, le nombre d'observations. Si R désigne l'étendue des données (c.-à-d., la différence absolue entre les observations extrêmes), chaque catégorie aura alors une largeur d'environ R/k. Par souci de simplicité, nous n'examinerons que les cas où les catégories ont toutes la même largeur.

Exemple 2.1 Organisation des données

Le Tableau 2.1 donne la teneur en humidité (%) de 50 échantillons indépendants de poudre de lait écrémé analysés en laboratoire.

TABLEAU 2.1: Teneur en humidité (%) de la poudre de lait écrémé

3.4	2.9	4.6	3.9	3.5	2.8	3.4	4.0	3.1	3.7
3.5	3.1	2.5	4.4	3.7	3.2	3.8	3.2	3.7	3.2
3.6	3.0	3.3	4.0	3.4	3.0	4.3	3.8	3.8	3.6
3.4	2.7	3.5	3.6	3.6	3.3	3.7	3.5	4.1	3.1
3.7	3.2	3.9	4.2	3.5	2.9	3.9	3.6	3.4	3.3

Le nombre de catégories devrait être d'environ $\sqrt{50}$, ou $k = 7$. L'étendue des données correspond à $R = 4.6 - 2.5 = 2.1$. Les catégories consécutives obtenues, chacune ayant une largeur de $R/k = 2.1/7 = 0.3$, sont 2.5 - 2.8, 2.8 - 3.1, . . ., 4.3 - 4.6. Par convention, toute observation qui tombe sur la limite de deux catégories est automatiquement rangée dans la catégorie supérieure. La table des fréquences pourrait prendre la forme du Tableau 2.2.

TABLEAU 2.2: Table des fréquences — Teneur en humidité (%) de la poudre de lait écrémé

Limites réelles de classe	Milieu de classe (X)	Fréquence de classe (f)	Fréquence cumulative
2.5 - 2.8	2.65	2	2
2.8 - 3.1	2.95	5	7
3.1 - 3.4	3.25	10	17
3.4 - 3.7	3.55	15	32
3.7 - 4.0	3.85	11	43
4.0 - 4.3	4.15	4	47
4.3 - 4.6	4.45	3	50

Les données ayant été mises en table des fréquences, on peut les présenter sous forme d'histogramme (Figure 2.1) en portant la fréquence de classe sur l'axe vertical et les limites réelles sur l'axe horizontal. On suppose alors que la valeur de toutes les observations d'une même catégorie est égale à celle du milieu de classe.

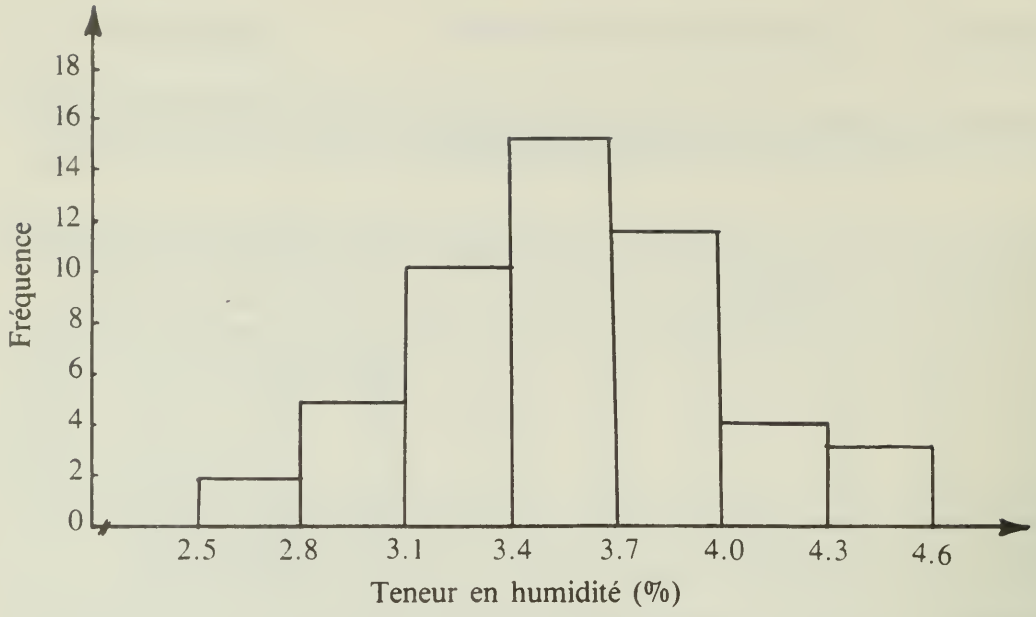


Figure 2.1 Histogramme représentant les données du Tableau 2.2

2.4 Mesures statistiques

Les grands ensembles de données, qu'ils soient composés de données brutes et non groupées ou de données résumées et groupées dans une table des fréquences, peuvent n'avoir aucune valeur informative. Il faut alors les caractériser, surtout si l'on veut s'en servir pour faire des comparaisons et prendre des décisions. Les caractéristiques de valeur centrale (ou centre) et les caractéristiques de dispersion (à partir du centre) sont des mesures statistiques typiques.

2.4.1 Moyenne

La mesure statistique de valeur centrale la plus répandue est la moyenne. Si l'on prend un échantillon d'observations $X_1, X_2, \ldots, X_n$, la **moyenne arithmétique de l'échantillon** (ou **moyenne de l'échantillon** ou, tout simplement, la **moyenne**) s'écrit $\overline{X}$ où

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}.$$

La **moyenne de la population** correspondante est désignée par la lettre μ (mu) de l'alphabet grec.

Dans le cas de données groupées, où les variables du milieu de classe $X_1, X_2, \ldots, X_k$ ont une fréquence respective de $f_1, f_2, \ldots, f_k$, la moyenne s'écrit:

$$\bar{X} = \frac{X_1 f_1 + X_2 f_2 + \ldots + X_k f_k}{f_1 + f_2 + \ldots + f_k} = \frac{\sum_{i=1}^{k} X_i f_i}{\sum_{i=1}^{k} f_i} .$$

2.4.2 Étendue

L'étendue d'un ensemble d'observations équivaut à la différence absolue entre l'observation la plus élevée et l'observation la moins élevée. Dans une table des fréquences, l'étendue est la différence absolue entre la limite réelle supérieure de la classe la plus élevée et la limite réelle inférieure de la classe la moins élevée.

2.4.3 Variance et écart-type

On peut estimer correctement le degré de variabilité d'un ensemble de données en calculant la mesure statistique de dispersion, que l'on appelle la variance. Dans un échantillon de n observations $X_1, X_2, \ldots, X_n$, la **variance de l'échantillon** s'écrit s^2 où

$$s^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n - 1} = \frac{\Sigma X_i^2 - \frac{(\Sigma X_i)^2}{n}}{n - 1} .$$

La **variance de la population** correspondante est représentée par le symbole σ^2 (sigma au carré), où le dénominateur $n - 1$ est remplacé par la taille de la population.

La racine carrée de la variance s'appelle **l'écart-type** que l'on désigne convenablement par s ou σ selon le cas. Dans une distribution de fréquences groupées, où un ensemble de variables de milieu de classe $X_1, X_2, \ldots, X_k$ ont une fréquence respective de $f_1, f_2, \ldots, f_k$, la variance s'écrit:

$$s^2 = \frac{\sum_{i=1}^{k} (X_i - \bar{X})^2 f_i}{\sum_{i=1}^{k} f_i - 1} = \frac{\Sigma X_i^2 f_i - \frac{(\Sigma X_i f_i)^2}{\Sigma f_i}}{\Sigma f_i - 1} .$$

2.4.4 Coefficient de variation

Il arrive souvent que l'on éprouve des difficultés à faire une comparaison entre deux ensembles ou plus de données exprimées dans des unités de mesure différentes. On se sert, aux fins de comparaison, d'une mesure appelée coefficient de variation et qui s'écrit:

$$\text{C.V.} = \frac{\text{écart-type}}{\text{moyenne}} = \frac{s}{\bar{X}} \times 100\% .$$

Le C.V., exprimé comme un pourcentage, permet de comparer la variabilité moyenne à la moyenne d'un ensemble de données, et ne comporte pas d'unité de mesure. Un petit pourcentage indique un groupe d'observations plutôt homogène.

Exemple 2.2 Mesures statistiques

À l'aide des données du Tableau 2.1, trouvez la moyenne, l'étendue, la variance, l'écart-type et le coefficient de variation de la teneur en humidité (%) des 50 échantillons de poudre de lait écrémé.

Dans le cas des données brutes,

$$\text{Moyenne} = \overline{X} = \frac{3.4 + 3.5 + \ldots + 3.3}{50} = \frac{175.5}{50} = 3.51$$

$$\text{Étendue} = 4.6 - 2.5 = 2.1$$

$$\Sigma X_i = 175.5, \ \Sigma X_i^2 = 625.37$$

$$\text{Variance} = s^2 = \frac{625.37 - \dfrac{(175.5)^2}{50}}{50 - 1} = 0.1911$$

$$\text{Écart-type} = s = \sqrt{0.1911} = 0.4372$$

$$\text{Coefficient de variation} = \text{C.V.} = \frac{s}{\overline{X}} \times 100\%$$

$$= \frac{0.4372}{3.51} \times 100\% = 12.46\%$$

Dans le cas des données groupées (se reporter au Tableau 2.2),

$$\text{Moyenne} = \overline{X} = \frac{(2 \times 2.65) + (5 \times 2.95) + \ldots + (3 \times 4.45)}{2 + 5 + \ldots + 3}$$

$$= \frac{178.1}{50} = 3.56$$

$$\text{Étendue} = 4.6 - 2.5 = 2.1$$

$$\Sigma X_i f_i = 178.1, \quad \Sigma X_i^2 f_i = 643.565$$

$$\text{Variance} = s^2 = \frac{643.565 - \dfrac{(178.1)^2}{50}}{50 - 1} = 0.1872$$

$$\text{Écart-type} = s = \sqrt{0.1872} = 0.4327$$

$$\text{Coefficient de variation} = \text{C.V.} = \frac{s}{\bar{X}} \times 100\%$$

$$= \frac{0.4327}{3.56} \times 100\% = 12.15\%$$

2.5 Probabilité

On utilise deux méthodes pour définir la probabilité: la **méthode classique** et celle dite de la **fréquence relative.** Dans l'optique classique, si un évènement donne lieu à n cas également possibles, dont r possèdent l'attribut A, la probabilité de A est égale à r/n. Cette définition est quelque peu restrictive; en effet, le calcul des probabilités suppose alors qu'on connaisse la valeur de n et qu'on sache avec certitude que tous les cas sont également possibles. Comme cela n'arrive pas souvent dans le cas des aliments et des produits agricoles, on a recours à une définition plus pragmatique qui fait appel au concept de la fréquence relative. Ainsi, on assimile la probabilité à la fréquence relative d'un grand nombre d'essais, c.-à-d., quand on parle de la probabilité d'un évènement A, on fait allusion à la fréquence relative de A dans un grand nombre d'essais analogues. Si l'évènement A se produit a fois en b essais et si le rapport a/b tend vers une limite, r/n, au fur et à mesure que n grandit arbitrairement, on dit que r/n est la probabilité de A et l'on écrit $P(A) = r/n$.

À la lumière de l'une ou l'autre de ces définitions, on constate que $P(A) \geqslant 0$, $P(A) \leqslant 1$, et, si $\overline{A}$ est le complément de A, on obtient $P(\overline{A}) = 1 - P(A)$. On remarquera que, si un évènement A est impossible, $r = 0$ et $P(A) = 0$. En revanche, si un évènement A est certain (c.-à-d., il se produit à chaque essai), $r = n$ et $P(A) = 1$.

Si, par exemple, on inspecte un échantillon de 200 oranges d'un gros envoi et qu'on constate que 10 d'entre elles sont atteintes d'une maladie, la proportion des oranges atteintes (ou la fréquence relative des oranges atteintes) est d'environ 10/200. La probabilité P qu'une orange soit atteinte sera donc de 0.05.

2.6 Permutations et combinaisons

La **permutation** est l'arrangement de r objets d'un ensemble de n objets dans un ordre défini. Le nombre de permutations différentes est représenté par ${}^{n}P_{r}$ et s'écrit:

$${}^{n}P_{r} = \frac{n!}{(n - r)!},$$

où $n! = n(n - 1)(n - 2) \ldots 1$ et $(n - r)! = (n - r)(n - r - 1) \ldots 1$, $0!$ étant égal à 1. Par exemple, $3! = 3 \cdot 2 \cdot 1 = 6$.

La **combinaison** est un sous-ensemble de r objets choisis au hasard parmi un ensemble de n objets. Le nombre de combinaisons différentes est représenté par $\binom{n}{r}$ et s'écrit:

$$\binom{n}{r} = \frac{n!}{r!\,(n - r)!}$$

Si l'on prend n essais répétés ayant chacun deux résultats possibles (défectueux et non défectueux, par exemple), le nombre total possible de séquences ou d'arrangements différents, ayant chacun x articles défectueux et $n - x$ articles non défectueux, est $\binom{n}{x}$.

2.7 Distributions statistiques

Tous les ensembles de données discrètes ou continues présentent un schéma de distribution. On peut plus facilement analyser un ensemble de données si l'on connaît les propriétés de la distribution à laquelle il appartient. Les trois principales distributions de probabilité dans le cas des données discrètes sont la **distribution binomiale,** la **distribution de Poisson** et la **distribution hypergéométrique.** Dans le cas des données continues, le calcul des probabilités se fait généralement à l'aide de la **distribution normale.**

2.7.1 Distribution binomiale

Si p représente la probabilité qu'un article soit défectueux, la probabilité que, dans un échantillon aléatoire de n objets, il y ait x articles défectueux est donnée par la distribution binomiale qui suit:

$$P\ (x \text{ articles défectueux parmi n objets}) = \binom{n}{x} p^x\,(1 - p)^{n - x}$$

où $x = 0, 1, 2, \ldots, n$.

Exemple 2.3 Distribution binomiale

Une technique d'étiquetage donne 20% d'éléments défectueux. Quelle est la probabilité de trouver deux éléments défectueux dans un échantillon de quatre?

Dans cet exemple, $n = 4$, $x = 2$, $p = 0.2$.

$$P\ (2 \text{ éléments défectueux}) = \binom{4}{2} (0.2)^2\,(0.8)^2$$
$$= 0.1536\ .$$

2.7.2 Distribution de Poisson

Dans le cas où l'évènement envisagé peut être qualifié de rare et aléatoire, n étant élevé et p peu élevé, le calcul des probabilités se fait à l'aide de la distribution de Poisson:

$$P\ (x \text{ éléments défectueux parmi n objets}) = \frac{e^{-\lambda}\ \lambda^x}{x!}$$

où $\lambda = np$ et $e = 2.718$. On peut généralement obtenir une approximation des probabilités binomiales à l'aide de la distribution de Poisson, si n est supérieur à 20 et p est inférieur à 0.05. Si n est supérieur à 100, p peut même atteindre 0.1.

Exemple 2.4 Distribution de Poisson

Reprenons l'exemple de l'étiquetage. Supposons que la proportion d'éléments défectueux soit de 5% et qu'on prélève à chaque heure un échantillon de quarante éléments. Quelle est la probabilité de trouver un élément défectueux dans un échantillon?

Dans cet exemple, n est élevé (40), p est peu élevé (0.05), et $\lambda = np = 2$. À l'aide de la distribution de Poisson, on obtient:

$$P\ (1 \text{ élément défectueux}) = \frac{e^{-2} \times 2^1}{1!} = 0.2707\ .$$

2.7.3 Distribution hypergéométrique

La distribution hypergéométrique sert à calculer les probabilités d'échantillons de petites populations. Si une population comprend N unités dont X sont défectueuses et qu'on tire au hasard un échantillon de n unités, la probabilité de trouver x unités défectueuses dans l'échantillon s'écrit:

$$P\ (x \text{ unités défectueuses}) = \frac{\binom{X}{x}\binom{N-X}{n-x}}{\binom{N}{n}}$$

où $x = 0, 1, 2, \ldots, n$.

Exemple 2.5 Distribution hypergéométrique

Si une population comprend vingt objets dont deux sont défectueux et on examine un échantillon de cinq objets, quelle est la probabilité que l'échantillon comporte un objet défectueux?

Dans cet exemple, X = 2, x = 1, N = 20 et n = 5.

$$P\ (1\ \text{objet défectueux}) = \frac{\binom{2}{1}\binom{18}{4}}{\binom{20}{5}} = 0.3947.$$

2.7.4 Distribution normale

La distribution qui s'utilise le plus souvent dans le cas des données continues est la distribution normale dont voici la formule:

$$P\ (X) = \frac{1}{\sigma\sqrt{2\pi}}\ e^{-1/2\{(X - \mu)/\sigma\}^2},\ \text{pour toutes les valeurs de X,}$$

où μ et σ correspondent respectivement à la moyenne et à l'écart-type de la population. La Figure 2.2 est une représentation graphique d'une distribution normale. La probabilité qu'une variable X, qui a une distribution normale de moyenne μ et d'écart-type σ, se situe entre deux valeurs a et b correspond à l'aire sous la courbe des fréquences, entre a et b.

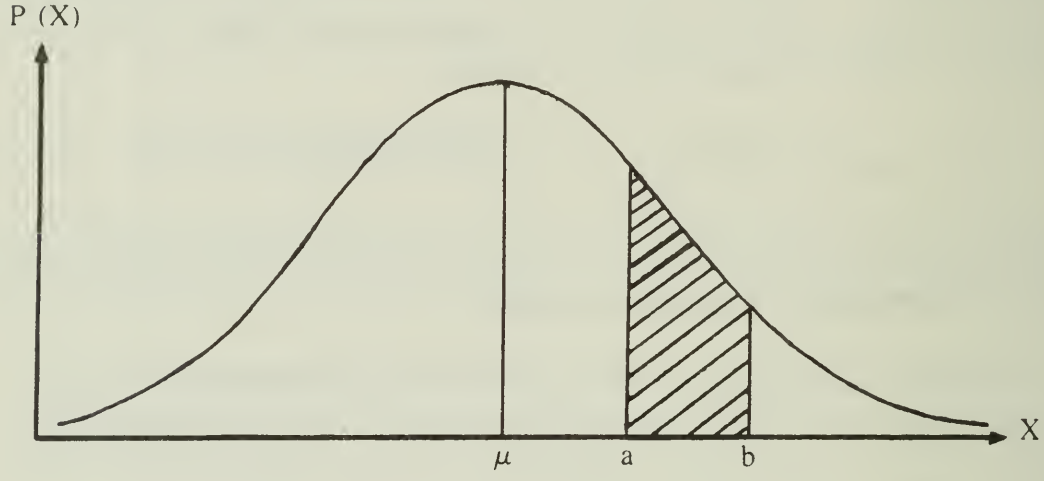


Figure 2.2 Une distribution normale type

La courbe normale des probabilités présente un certain nombre de caractéristiques importantes:

1. μ et σ étant définis, la courbe des probabilités est entièrement déterminée.
2. La courbe est symétrique par rapport à un axe vertical qui passe par la moyenne. Les observations tendent à se grouper autour de la moyenne.
3. L'aire totale sous la courbe est égale à 1.
4. Bien que la courbe s'étende à l'infini dans les deux sens, elle est à toutes fins utiles négligeable au-delà de 3σ de chaque côté de la moyenne. On sait, par ailleurs, empiriquement que (Figure 2.3):

68.26% de l'aire se trouve entre $\mu \pm 1\sigma$,

95.44% de l'aire se trouve entre $\mu \pm 2\sigma$,

99.73% de l'aire se trouve entre $\mu \pm 3\sigma$.

La courbe normale des probabilités étant entièrement dépendante de la valeur de μ et de σ, sa forme et son emplacement varient en conséquence, ce qui entraîne d'infinies possibilités de distributions. Vouloir établir des tableaux distincts des probabilités normales ou des aires correspondant à toutes les valeurs imaginables de μ et de σ constituerait un véritable travail de bénédictin. Heureusement, cela n'est pas nécessaire. On peut en effet transformer la variable aléatoire X normalement distribuée, possédant une moyenne μ et un écart-type σ, en une nouvelle variable aléatoire appelée Z qui s'écrit:

$$Z = \frac{X - \mu}{\sigma}$$

et qui a une **distribution normale type** où $\mu = 0$ et $\sigma = 1$. La courbe qui correspond à la distribution normale type a son centre au point zéro et la majeure partie de son aire entre -3 et 3. La probabilité que Z se trouve entre deux nombres a et b (on suppose que $a < b$) est donnée dans le Tableau 1 de l'Appendice.

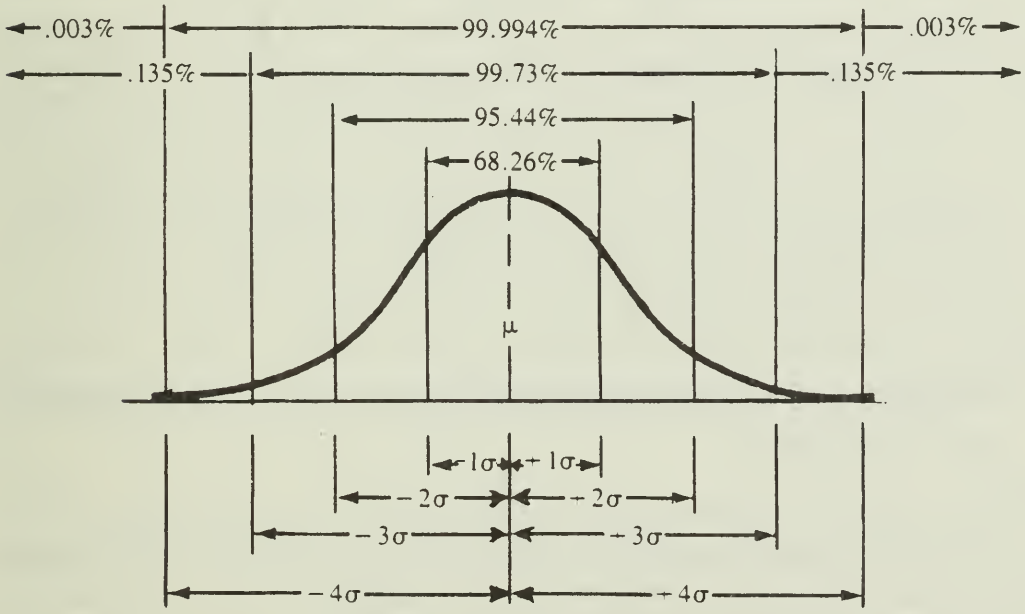


Figure 2.3 Pourcentages de la distribution normale

Exemple 2.6 Distribution normale

Si l'on suppose que la teneur en humidité de la poudre de lait écrémé présente une distribution normale, la moyenne μ étant égale à 3.51 (%) et l'écart-type σ, à 0.4372 (%), quelle est la probabilité que la teneur en humidité d'un échantillon de poudre de lait écrémé choisi au hasard se situe entre 2.9 et 3.8 (%)?

$$\text{Lorsque } X = 2.9,\ Z = \frac{2.9 - \mu}{\sigma} = \frac{2.9 - 3.51}{0.4372} = -1.40 .$$

$$\text{Lorsque } X = 3.8,\ Z = \frac{3.8 - \mu}{\sigma} = \frac{3.8 - 3.51}{0.4372} = 0.66 .$$

La probabilité que la teneur en humidité (X) se situe entre 2.9 et 3.8 est la même que la probabilité ou l'aire sous la courbe normale type (Z) entre les points -1.40 et 0.66, c.-à-d., $P(-1.40 < Z < 0.66)$, comme on peut le constater dans la Figure 2.4. Ainsi, à la lecture du Tableau 1 de l'Appendice, on obtient:

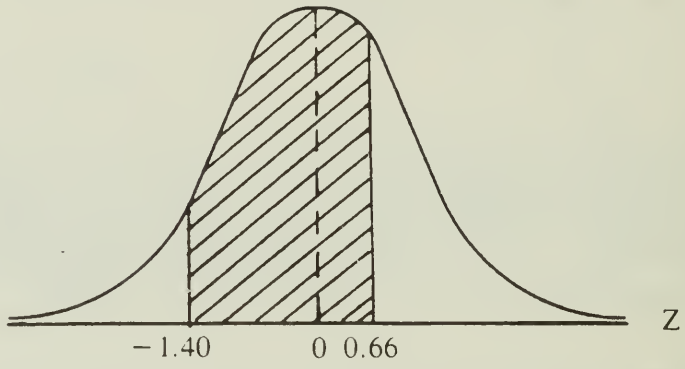


Figure 2.4 Distribution normale type de l'exemple 2.6

$$\begin{aligned} P(-1.40 < Z < 0.66) &= P(-1.40 < Z < 0) + P(0 < Z < 0.66) \\ &= P(0 < Z < 1.40) + P(0 < Z < 0.66) \\ &= 0.4192 + 0.2454 \\ &= 0.6646 . \end{aligned}$$

2.8 Distribution de la moyenne de l'échantillon et de la variance de l'échantillon

Si une population compte N éléments, le nombre total d'échantillons distincts de taille n que l'on peut prélever est $\binom{N}{n}$. Chacun de ces échantillons a une moyenne $\overline{X}$ et ces $\overline{X}$ ont leur propre distribution. Si la population de départ a une distribution normale, la variable $\overline{X}$ aura elle aussi une distribution normale. Même si la population de départ n'est pas normale, un théorème très important en statistique mathématique, appelé le théorème de la limite centrale, nous apprend que la distribution de $\overline{X}$ est approximativement normale et que cette approximation augmente au fur et à mesure que n croît. Si μ et σ représentent respectivement la moyenne et l'écart-type de la population de départ, la moyenne et l'écart-type de la distribution d'échantillonnage de $\overline{X}$, que l'on désigne par $\mu_{\overline{X}}$ et $\sigma_{\overline{X}}$, s'écrivent:

$$\mu_{\overline{X}} = \mu \quad \text{et} \quad \sigma_{\overline{X}} = \sigma \sqrt{\frac{N - n}{n(N - 1)}} .$$

Dans les grandes populations (valeurs N élevées), $\sigma_{\bar{X}}$ est sensiblement égal à $\sigma/\sqrt{n}$ et représente **l'erreur-type** de la moyenne. On remarquera que l'écart-type désigne la variation moyenne des observations d'unités individuelles, alors que l'erreur-type représente la variation aléatoire d'une estimation tirée d'une expérience complète. Si une variable a une distribution normale quelconque, on peut la réduire à une distribution normale type en soustrayant la moyenne et en la divisant par son écart-type. Comme la distribution de $\bar{X}$ est normale, la moyenne étant μ et l'écart-type, $\sigma/\sqrt{n}$, on peut réduire la variable en soustrayant μ et en la divisant par $\sigma/\sqrt{n}$. On obtient ainsi:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} .$$

2.9 Distribution de t

La principale difficulté qu'on rencontre est que σ est généralement inconnu et qu'il faut l'estimer à l'aide de l'écart-type de l'échantillon s. Toutefois, par soustraction du μ et division par $s/\sqrt{n}$, on n'obtient plus une distribution normale type de la variable $\dfrac{\bar{X} - \mu}{s/\sqrt{n}}$, mais plutôt une **distribution de t.** Tout comme la distribution normale type, la distribution de t est symétrique, en forme de cloche et centrée à zéro. Il n'y a pas cependant qu'une seule distribution de t, mais une véritable famille de distributions dont chaque membre est caractérisé par son **degré de liberté (d.f.), f,** que l'on définit par $f = n - 1$.

Alors que dans la distribution normale type, 95% de l'aire se situe entre $\pm$ 1.96 (se reporter au Tableau 1 de l'Appendice), il y a, pour chaque distribution de t, un nombre, désigné $t_{.025}$, tel que 95% de l'aire sous la courbe t se trouve entre $-t_{.025}$ et $+t_{.025}$. Les valeurs réelles de $t_{.025}$ varient en fonction des degrés de liberté; ces valeurs, ainsi que d'autres valeurs de t, sont indiquées en Appendice (Tableau 2). Trouvons, par exemple, les deux nombres de la distribution de t entre lesquels se situe centralement 95% de l'aire lorsque $n = 15$.

Pour $n = 15$, $f = 14$ et, par conséquent, $t_{.025} = 2.1448$ suivant le Tableau 2 de l'Appendice. Ainsi, la distribution de t pour 14 degrés de liberté se situe dans une proportion de 95% entre -2.1448 et 2.1448. On remarquera que, au fur et à mesure que n augmente, les valeurs de $t_{.025}$ et $-t_{.025}$ se rapprochent. Quand n est très élevé, $t_{.025} = 1.96$, soit la valeur comparable de la distribution normale type. À toutes fins utiles, la distribution de t correspond à la distribution normale type lorsque $n > 30$.

2.10 Distribution de chi carré

Tout comme il y a une moyenne $\bar{X}$ pour chacun des $\binom{N}{n}$ échantillons de taille n qu'on peut tirer d'une population qui compte N éléments, chaque échantillon possède également un écart-type s et une variance s^2. De la même manière que $\bar{X}$ a une distribution normale (ou presque normale), s^2 a aussi une distribution.

On utilise habituellement la quantité $(n - 1)s^2/\sigma^2$, qui a ce qu'il est convenu d'appeler une distribution de **chi carré** (χ^2).

Pour calculer les probabilités de $(n - 1)s^2/\sigma^2$ (et, de ce fait, celles de s^2), on doit trouver les aires appropriées sous la distribution de χ^2, dont les valeurs sont indiquées dans le Tableau 3 de l'Appendice. Tout comme la distribution de t, la distribution de χ^2 constitue une famille de distributions caractérisées dans chaque cas par un nombre de degrés de liberté, $f = n - 1$. Pour trouver les deux valeurs de la distribution de χ^2 entre lesquelles se situe centralement 95% de l'aire, on désigne la valeur inférieure par $\chi^2_{.975}$ et la valeur supérieure par $\chi^2_{.025}$ et on obtient alors les valeurs correspondantes dans le Tableau 3 de l'Appendice. À l'aide de ce tableau, par exemple, on constate que, si $n = 10$, les valeurs de $\chi^2_{.975}$ et $\chi^2_{.025}$ pour $f = n - 1 = 9$ degrés de liberté sont respectivement de 2.700 et de 19.023.

2.11 Estimation: Intervalles de confiance

Les **estimateurs ponctuels** de la moyenne et de la variance de la population désignent la moyenne $\overline{X}$ et la variance s^2 de l'échantillon, calculées à partir des observations dans l'échantillon pour estimer les paramètres correspondants μ et σ^2 de la population. Toutefois, les statistiques de l'échantillon n'auront vraisemblablement pas la même valeur que les paramètres auxquels elles se rapportent dans la population. D'une façon générale, il est donc souhaitable d'établir un intervalle à l'intérieur duquel on peut s'attendre, en se fondant sur un certain degré de confiance, à trouver un paramètre de la population. Cette méthode, appelée **intervalle d'estimation,** offre un **intervalle de confiance** qui sert à cerner la valeur réelle d'un paramètre d'une population en tenant compte du caractère incertain des estimations de l'échantillon. Le **niveau de confiance** est représenté par $1 - \alpha$, où la lettre grecque α (alpha) désigne le **niveau de signification.** Par exemple, si $\alpha = 0.05$, l'intervalle de confiance calculé portera un niveau de confiance de 0.95 ou 95%.

1. **Intervalle de confiance de la moyenne μ d'une distribution normale**

   (i) **Variance, σ^2, connue**

   L'intervalle de confiance général de $100(1 - \alpha)\%$ pour μ s'écrit:

$$\left(\overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \;,\; \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right).$$

   Plus précisément, un intervalle de confiance de 95% pour μ s'écrit:

$$\left(\overline{X} - 1.96\frac{\sigma}{\sqrt{n}} \;.\; \overline{X} + 1.96\frac{\sigma}{\sqrt{n}}\right).$$

Dans cet exemple, $1 - \alpha = 0.95$ et $Z_{\alpha/2} = Z_{.025} = 1.96$, soit la valeur tirée du Tableau 1 de l'Appendice, de sorte que 95% de l'aire sous la courbe normale type se situe entre $-Z_{.025}$ et $+Z_{.025}$. On peut être sûr à 95% que cet intervalle contient la moyenne μ.

(ii) **Variance, σ^2, inconnue**

Quand la variance σ^2 est inconnue et qu'elle est estimée à l'aide de la variance de l'échantillon s^2, l'intervalle de confiance de $100(1 - \alpha)\%$ pour la moyenne μ s'écrit:

$$\left(\bar{X} - t_{\alpha/2} \frac{s}{\sqrt{n}} \ , \ \bar{X} + t_{\alpha/2} \frac{s}{\sqrt{n}}\right) ,$$

où $t_{\alpha/2}$ correspond à la valeur de t tirée du Tableau 2 de l'Appendice pour $(n - 1)$ degrés de liberté. Ainsi, dans le cas d'un intervalle de confiance de 95%, on peut lire dans la colonne sous $t_{.025}$ les valeurs de t correspondant à $(n - 1)$ degrés de liberté.

2. Intervalle de confiance de la variance σ^2 d'une distribution normale

Pour déterminer un intervalle de confiance de 95% correspondant à une variance σ^2, on doit calculer s^2, la variance de l'échantillon établie à partir d'un échantillon de n éléments choisis au hasard, et lire dans le Tableau 3 de l'Appendice les valeurs de $\chi^2_{.975}$ et de $\chi^2_{.025}$ pour $(n - 1)$ degrés de liberté. Par conséquent, un intervalle de confiance de 95% pour une variance σ^2 s'écrit:

$$\left(\frac{(n - 1)\ s^2}{\chi^2_{.025}} \ , \ \frac{(n - 1)\ s^2}{\chi^2_{.975}} \right) .$$

Un intervalle de confiance plus général de $100(1 - \alpha)\%$ pour la même variance σ^2 s'écrit:

$$\left(\frac{(n - 1)\ s^2}{\chi^2_{\alpha/2}} \ , \ \frac{(n - 1)\ s^2}{\chi^2_{1-\alpha/2}} \right) ,$$

où $\chi^2_{\alpha/2}$ et $\chi^2_{1-\alpha/2}$ représentent les valeurs de χ^2 indiquées dans le Tableau 3 de l'Appendice pour $(n - 1)$ degrés de liberté.

On peut donc exprimer un intervalle de confiance de 95% pour l'écart-type σ d'une distribution normale de la façon suivante:

$$\left(\sqrt{\frac{(n - 1)\ s^2}{\chi^2_{.025}}} \ , \ \sqrt{\frac{(n - 1)\ s^2}{\chi^2_{.975}}} \right) .$$

3. Intervalle de confiance d'une proportion

Si une variable aléatoire discrète a une distribution binomiale, il se peut qu'on veuille estimer la proportion p d'éléments défectueux au sein de la population. Supposons qu'on tire un échantillon aléatoire de taille n et qu'on trouve X éléments défectueux, X/n étant la proportion d'éléments défectueux dans l'échantillon. Si n est élevé et que p n'est pas trop proche de 0 ou de 1, le théorème de la limite centrale nous permet d'utiliser l'approximation normale de la distribution binomiale, ce qui signifie que

$$\frac{X/n - p}{\sqrt{\frac{p(1-p)}{n}}}$$

suit approximativement une distribution normale type. L'intervalle de confiance de 95% de p s'exprime donc ainsi:

$$\left(\frac{X}{n} - 1.96\sqrt{\frac{p\,(1-p)}{n}} \;,\; \frac{X}{n} + 1.96\sqrt{\frac{p\,(1-p)}{n}} \right) .$$

Malheureusement, cet intervalle dépend de p qui, évidemment, est inconnu. Toutefois, si nous remplaçons p par l'estimation ponctuelle X/n, nous obtenons l'intervalle de confiance approximatif (95%) suivant:

$$\left[\frac{X}{n} - 1.96\sqrt{\frac{\frac{X}{n}\left(1 - \frac{X}{n}\right)}{n}} \;,\; \frac{X}{n} + 1.96\sqrt{\frac{\frac{X}{n}\left(1 - \frac{X}{n}\right)}{n}} \right]$$

Exemple 2.7 Intervalles de confiance de la moyenne et de la variance

Calculer les intervalles de confiance de 95% pour μ, σ^2 et σ à l'aide des données sur la teneur en humidité de la poudre de lait écrémé présentées dans la Section 2.3, ainsi que des calculs qui en découlent à l'Exemple 2.2.

Dans le cas présent, $\overline{X} = 3.51$, $s^2 = 0.1911$, $s = 0.4372$ et $n = 50$.

L'intervalle de confiance de 95% pour μ correspond à:

$$\left(3.51 - \frac{(2.01)\ (0.4372)}{\sqrt{50}} \;,\; 3.51 + \frac{(2.01)\ (0.4372)}{\sqrt{50}} \right)$$

soit (3.39 , 3.63), où la valeur de $t_{.025}$ est sensiblement égale à 2.01 pour 49 degrés de liberté (Tableau 2 de l'Appendice). Nous sommes donc presque sûrs, c.-à-d., à 95%, que la véritable teneur moyenne en humidité de la poudre de lait écrémé se situe entre 3.39 et 3.63 (%).

Pour obtenir les intervalles de confiance de 95% de σ^2 et de σ, il faut d'abord tirer du Tableau 3 de l'Appendice les valeurs de χ^2 requises pour 49 degrés de liberté:

$$\chi^2_{.025} = 70.222 \text{ et } \chi^2_{.975} = 31.555.$$

Ainsi, l'intervalle de confiance de 95% de σ^2 correspond à:

$$\left(\frac{(50 - 1)\,(0.1911)}{70.222} \;,\; \frac{(50 - 1)\,(0.1911)}{31.555} \right)$$

soit (0.1333 , 0.2967).

L'intervalle de confiance de 95% de σ est donc:

$$(\sqrt{0.1333} \;,\; \sqrt{0.2967}) = (0.3651 \;,\; 0.5447).$$

Exemple 2.8 Intervalle de confiance d'une proportion

L'inspection de 100 pommes prélevées au hasard d'un gros lot de pommes permet d'y trouver 15 mauvaises pommes. Calculer l'intervalle de confiance de 95% de p, la proportion véritable de mauvaises pommes dans l'ensemble du lot.

Dans cet exemple,

$$n = 100,\ X = 15,\ \frac{X}{n} \doteq 0.15 \text{ et } \sqrt{\frac{\frac{X}{n}\left(1 - \frac{X}{n}\right)}{n}} = 0.0357.$$

On calcule alors l'intervalle de confiance approximatif (95%) de p comme suit:

$$(0.15 - 1.96 \times 0.0357 \;,\; 0.15 + 1.96 \times 0.0357) = (0.08 \;,\; 0.22).$$

On est donc sûr à 95% que la proportion véritable de mauvaises pommes dans le lot se situe entre 8% et 22%.

2.12 Vérification d'hypothèse

Alors que l'estimation statistique se fonde sur les observations dans l'échantillon pour estimer des paramètres inconnus au moyen de valeurs uniques ou d'intervalles, la vérification d'hypothèse permet de tester la validité de certaines hypothèses que l'on formule à l'égard de ces paramètres.

Hypothèses nulles et alternatives

L'hypothèse nulle représente l'hypothèse principale à vérifier et s'écrit H_0. Toutes les autres hypothèses complémentaires sont des **hypothèses alternatives** que l'on désigne par H_A. Par exemple, dans le cadre du contrôle du poids de remplissage des pêches mises en conserve sur une chaîne de production, on pourrait formuler l'hypothèse que le poids de remplissage moyen est supérieur à 32 onces. On obtient donc H_0: $\mu = 32$ et H_A: $\mu > 32$. S'il y a deux chaînes de production qui effectuent le même travail, on pourrait supposer que le poids de remplissage moyen est identique pour les deux. On obtient dans ce cas H_0: $\mu_1 = \mu_2$ et H_A: $\mu_1 \neq \mu_2$, où μ_1 et μ_2 désignent le poids de remplissage véritable des pêches mises en conserve sur les chaînes de production 1 et 2 respectivement.

Erreurs de type I et de type II

La vérification d'hypothèse peut donner lieu à deux types d'erreur. Une **erreur de type I** correspond au rejet de H_0 alors qu'en réalité l'hypothèse est vraie. La probabilité de commettre une erreur de type I s'appelle le **niveau de signification** du test qu'on désigne par la lettre grecque α (alpha). On commet une **erreur de type II** lorsqu'on accepte H_0 alors que dans les faits l'hypothèse est fausse. La probabilité d'une erreur de type II est représentée par la lettre grecque ß (bêta) et 1-ß désigne la **puissance** du test. Comme il est généralement difficile de prévoir la probabilité de commettre une erreur de type II, on a élaboré des procédures d'essai qui conviennent à l'erreur de type I et au moyen desquelles on peut exercer un contrôle sur celle-ci.

On prend la décision d'accepter ou de rejeter l'hypothèse nulle en établissant des **régions d'acceptation** et des **régions critiques (de rejet)** en fonction d'un **niveau de confiance** de $1 - \alpha$. Les **valeurs critiques** servent de limites pour séparer la région d'acceptation de la région critique. Si l'on se reporte au Tableau 1 de l'Appendice, les valeurs critiques d'une distribution normale type dans le cas de $\alpha = 0.05$ sont $-Z_{\alpha/2} = -1.96$ et $Z_{\alpha/2} = 1.96$. Si, d'après nos calculs, la statistique d'essai Z se situe dans la région critique, définie par les valeurs plus petites que -1.96 ou plus grandes que 1.96, l'hypothèse nulle est rejetée suivant un niveau de signification de 5%. L'hypothèse nulle sera acceptée si Z se situe dans la région d'acceptation délimitée par -1.96 et 1.96 (se reporter à la Figure 2.5).

Tests unilatéraux et bilatéraux

Les régions critiques d'une distribution peuvent varier d'un test à l'autre, même si le niveau de signification demeure le même, en fonction de l'hypothèse alternative énoncée. La procédure d'essai de n'importe quelle hypothèse statistique liée à une hypothèse alternative unilatérale, comme

$$H_0: \mu = 4,\ H_A: \mu > 4 \text{ ou } H_0: \mu = 4,\ H_A: \mu < 4\ ,$$

s'appelle un **test unilatéral.** Toute la région critique de l'hypothèse alternative $H_A: \mu > 4$ se situe à l'extrême droite de la distribution, alors que toute la région critique de l'hypothèse alternative $H_A: \mu < 4$ se trouve à l'extrémité gauche. Dans tels cas, où une distribution symétrique y a lieu, si $\alpha = 0.05$ ou 5%, la région d'acceptation doit comprendre toute l'aire sous la courbe d'un côté de la moyenne ainsi que 45% de l'aire, à partir de la moyenne, de l'autre côté (se reporter à la Figure 2.5).

La procédure d'essai de n'importe quelle hypothèse statistique par rapport à une hypothèse alternative bilatérale, comme

$$H_0: \mu = 4,\ H_A: \mu \neq 4\ ,$$

s'appelle un **test bilatéral.** L'hypothèse alternative indique que $\mu < 4$ ou $\mu > 4$. La région critique occupe donc une aire égale aux deux extrémités de la distribution. Dans le cas d'une distribution symétrique dans l'exemple présent, si $\alpha = 0.05$ ou 5%, la région d'acceptation doit être divisée également de part et d'autre de la moyenne, c.-à-d., occuper une aire de 47.5% de chaque côté de celle-ci (se reporter à la Figure 2.5).

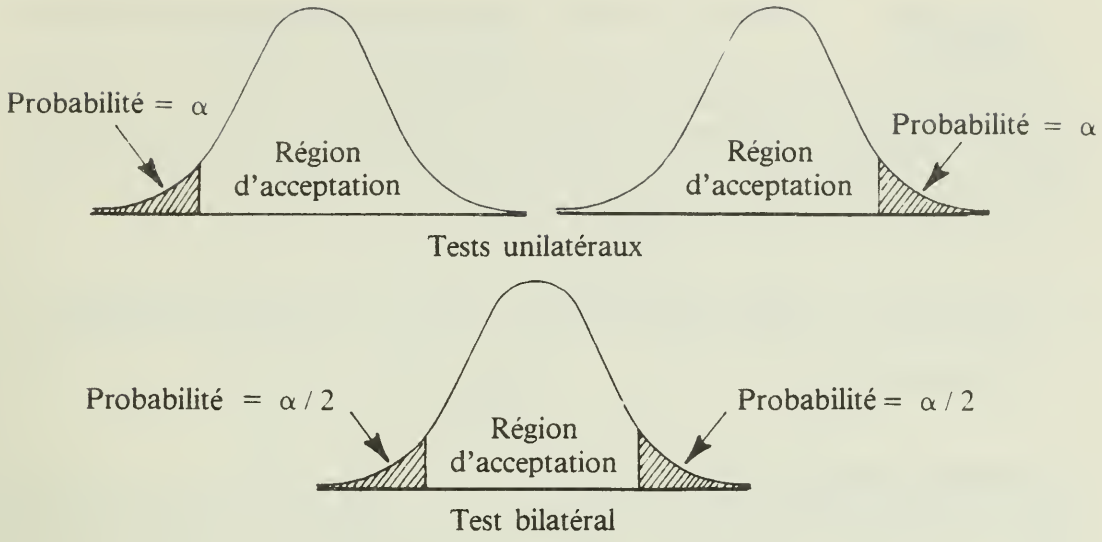


Figure 2.5 Régions d'acceptation et régions critiques des tests unilatéraux et bilatéraux dans le cas d'une distribution normale

2.12.1 Tests de signification

1. **Vérification d'une moyenne de valeur μ_0, variance σ^2 connue: test de Z**

Hypothèse nulle $H_0: \mu = \mu_0$

Statistique d'essai $Z = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$

Hypothèse alternative	**Rejet de H_0 selon un niveau de signification de 0.05 si**
$\mu \neq \mu_0$	$Z > 1.96$ ou $Z < -1.96$
$\mu > \mu_0$	$Z > 1.645$
$\mu < \mu_0$	$Z < -1.645$

2. **Vérification d'une moyenne de valeur μ_0, variance σ^2 inconnue: test de t**

Hypothèse nulle $H_0: \mu = \mu_0$

Statistique d'essai $t = \dfrac{\bar{X} - \mu_0}{s/\sqrt{n}}$, avec $f = (n - 1)$ d.f.

Hypothèse alternative	Rejet de H_0 selon un niveau de signification de 0.05 si
$\mu \neq \mu_0$	$t > t_{.025}$ ou $t < -t_{.025}$
$\mu > \mu_0$	$t > t_{.05}$
$\mu < \mu_0$	$t < -t_{.05}$

3. **Vérification des différences entre deux moyennes: variances inconnues**

Hypothèse nulle $H_0: \mu_1 = \mu_2$ ou $H_0: \mu_1 - \mu_2 = 0$

Statistique d'essai

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_p \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}},$$

avec $f = (n_1 + n_2 - 2)$ d.f.

où $s_p^2 = \dfrac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}{n_1 + n_2 - 2}$.

Hypothèse alternative	Rejet de H_0 selon un niveau de signification de 0.05 si
$\mu_1 \neq \mu_2$	$t > t_{.025}$ ou $t < -t_{.025}$
$\mu_1 > \mu_2$	$t > t_{.05}$
$\mu_1 < \mu_2$	$t < -t_{.05}$

4. **Vérification d'une proportion de valeur p_0**

Hypothèse nulle $H_0: p = p_0$

Statistique d'essai

$$Z = \frac{\frac{X}{n} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}$$

où X correspond au nombre d'occurrences de l'attribut dans les n essais.

Hypothèse alternative	Rejet de H_0 selon un niveau de signification de 0.05 si
$p \neq p_0$	$Z > 1.96$ ou $Z < -1.96$
$p > p_0$	$Z > 1.645$
$p < p_0$	$Z < -1.645$

Exemple 2.9 Vérification d'une valeur moyenne

Supposons que le poids moyen des oeufs pondus par des poules nourries d'une façon particulière est supérieur à 55 grammes. Pour vérifier cette hypothèse, on choisit 25 oeufs au hasard et on pèse chacun d'eux, ce qui donne $\overline{X} = 56.0$ g et $s = 6.0$ g. À l'aide d'un test unilatéral selon un niveau de signification de 5%, vérifiez l'hypothèse H_0: $\mu = 55$ par rapport à H_A: $\mu > 55$.

La statistique d'essai appropriée se lit comme suit:

$$t = \frac{\overline{X} - \mu_0}{s/\sqrt{n}} = \frac{56.0 - 55}{6.0/\sqrt{25}} = 0.83 .$$

La valeur de $t_{.05}$ pour $f = n - 1 = 24$ degrés de liberté est de 1.71 (Tableau 2 de l'Appendice). Étant donné que la valeur de t calculée est inférieure à la valeur de t du tableau, nous ne pouvons pas rejeter l'hypothèse nulle selon le niveau de signification de 5%. La preuve n'est donc pas suffisante pour que nous puissions conclure que cette nourriture spéciale permet aux poules de pondre des oeufs dont le poids moyen est supérieur à 55 grammes.

Exemple 2.10 Vérification de la différence entre deux moyennes

Afin de comparer l'effet moyen de deux traitements différents sur une production de maïs, on applique le traitement 1 à 25 champs de maïs et le traitement 2 à 20 autres champs semblables. On obtient les moyennes et les variances correspondantes du nombre de boisseaux de maïs moissonné dans chaque champ de la façon suivante:

$$\overline{X}_1 = 83, \ \overline{X}_2 = 64, \ s_1^2 = 7.3, \ s_2^2 = 9.8 \ .$$

Vérifiez l'hypothèse H_0: $\mu_1 = \mu_2$ par rapport à H_A: $\mu_1 \neq \mu_2$, selon un niveau de signification de 0.05. La statistique d'essai dans ce cas s'écrit:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{83 - 64}{\sqrt{8.4 \left(\frac{1}{25} + \frac{1}{20} \right)}} = 21.9$$

$$\text{où } s_p^2 = \frac{(25 - 1)(7.3) + (20 - 1)(9.8)}{25 + 20 - 2} = 8.4 .$$

Suivant le Tableau 2 de l'Appendice, la valeur de $t_{.025}$ pour $f = n_1 + n_2 - 2 = 43$ degrés de liberté est environ 2.02. Étant donné que la valeur de t calculée (21.9) est supérieure à la valeur de t du tableau (2.02), nous rejetons l'hypothèse nulle et concluons que, selon le niveau de signification de 5%, la différence entre la production moyenne de maïs des deux traitements est statistiquement significative.

Exemple 2.11 Vérification de la valeur d'une proportion

À l'aide de l'échantillon aléatoire de 100 pommes qui ont fait l'objet d'une inspection dans l'Exemple 2.8, examinez l'hypothèse selon laquelle, pour un niveau de signification de 5%, la proportion de mauvaises pommes dans le lot correspondant est de 12% tout au plus.

Nous devons dans ce cas vérifier l'hypothèse H_0: $p = 0.12$ par rapport à H_A: $p > 0.12$. Le calcul de la statistique d'essai applicable est le suivant:

$$Z = \frac{\frac{X}{n} - p_0}{\sqrt{\frac{p_0 (1 - p_0)}{n}}} = \frac{0.15 - 0.12}{\sqrt{\frac{0.12 \times 0.88}{100}}} = 0.923 .$$

Puisque cette valeur de Z calculée (0.923) est inférieure à 1.645, les données ne permettent pas de justifier le rejet de l'hypothèse nulle et, par conséquent, on conclut qu'à un niveau de signification de 5% tout au plus 12% des pommes dans ce lot sont mauvaises.

De plus, on remarquera en se reportant à l'Exemple 2.8 que $p_0 = 0.12$ ou 12% se situe dans l'intervalle de confiance de 95% évalué pour p, ce qui vient confirmer nos résultats ci-dessus.

2.13 Régression linéaire et corrélation

2.13.1 Régression linéaire

L'analyse de corrélation mesure le degré d'association entre deux variables quantitatives alors que l'analyse de régression cherche à déterminer plus précisément la nature de cette relation à des fins de prévision. Pour prévoir le comportement de deux variables présentant une relation linéaire, on se sert d'une équation de régression linéaire $Y = A + BX$, où Y représente la variable dépendante inconnue et X, la variable indépendante connue. La constante A, appelée la **coordonnée à l'origine Y,** correspond à la valeur de Y au point où la ligne coupe l'axe des Y. La constante B, appelée le **coefficient de régression,** correspond à la **pente** de la ligne et représente le changement subi par la variable Y suite à un changement d'unité de mesure dans la valeur de X.

Dans le cas de tout échantillon de n observations Y_i correspondant à certaines (n) valeurs X_i, on obtient la valeur prévue de Y pour n'importe quelle valeur fixe de X, que l'on rend par $\hat{Y}$ (lu Y-accent circonflexe), en tirant la **droite de meilleur ajustement** pour ces données, à l'aide de la **méthode des moindres carrés** qui donne **l'équation de régression (estimée)** suivante:

$$\hat{Y} = a + bX$$

$$\text{où } b = \frac{\Sigma XY - \frac{(\Sigma X)\,(\Sigma Y)}{n}}{\Sigma X^2 - \frac{(\Sigma X)^2}{n}} \quad \text{et} \quad a = \frac{\Sigma Y}{n} - b\left(\frac{\Sigma X}{n}\right).$$

Ainsi, pour n'importe quelle valeur de X spécifiée, comme $X = X_0$, la valeur prévue correspondante de Y s'écrit $\hat{Y} = a + bX_0$. On peut calculer un **intervalle de prévision** de 95% associé à une estimation d'intervalle de Y comme suit:

$$\hat{Y} \pm t_{.025}\,(s_{\hat{y}}) \quad ,$$

où $t_{.025}$ correspond à la valeur de t du Tableau 2 de l'Appendice pour $(n - 2)$ degrés de liberté et $s_{\hat{y}}$ à l'erreur-type (S.E.) d'une seule valeur prévue pour Y, soit

$$s_{\hat{y}}^2 = s_{y.x}^2 \left(1 + \frac{1}{n} + \frac{(X_0 - \overline{X})^2}{\Sigma X^2 - \frac{(\Sigma X)^2}{n}}\right) ,$$

où $s_{y.x}$ représente l'erreur-type d'estimation d'une régression linéaire de Y sur X et

$$s_{y.x}^2 = \frac{\Sigma(Y - \hat{Y})^2}{n - 2} = \frac{\Sigma Y^2 - a(\Sigma Y) - b(\Sigma XY)}{n - 2} .$$

Par conséquent, l'intervalle de prévision de 95% de Y peut aussi s'écrire:

$$\hat{Y} \pm t_{.025}\sqrt{s_{y.x}^2 \left(1 + \frac{1}{n} + \frac{(X_0 - \overline{X})^2}{\Sigma X^2 - \frac{(\Sigma X)^2}{n}}\right)} .$$

2.13.2 Corrélation

On mesure l'association linéaire entre deux variables à l'aide d'un coefficient de corrélation calculé, pour un échantillon de n observations, de la façon suivante:

$$r = \frac{\Sigma XY - \frac{(\Sigma X)\,(\Sigma Y)}{n}}{\sqrt{\left(\Sigma X^2 - \frac{(\Sigma X)^2}{n}\right)\left(\Sigma Y^2 - \frac{(\Sigma Y)^2}{n}\right)}}$$

La valeur de r peut se situer n'importe où entre -1 et $+1$. Les valeurs négatives de r indiquent une relation inverse entre X et Y, alors qu'on obtient des valeurs positives de r lorsqu'il y a une relation directe entre les deux variables. Si $r = 0$, il n'y a aucune relation linéaire entre les variables X et Y.

Il est également utile de connaître le degré de la relation linéaire entre X et Y, d'après la grandeur absolue du coefficient de corrélation. À cette fin, on calcule la valeur de r^2, soit le **coefficient de détermination,** qui mesure la proportion de la variation totale de Y attribuable à l'association linéaire entre X et Y.

Pour vérifier la signification de r, on fixe à zéro, c.-à-d., H_0: $\rho = 0$, le coefficient de corrélation de la population, représenté par ρ (rho). On évalue ensuite ainsi la statistique d'essai

$$t = \frac{r}{\sqrt{\frac{1 - r^2}{n - 2}}}$$

qui a une distribution de t munie de $(n - 2)$ degrés de liberté. On prend la décision d'accepter ou de rejeter H_0 comme on l'a fait précédemment, en comparant les valeurs calculées et les valeurs de t du tableau selon un niveau de signification donné.

Exemple 2.12 Analyse de régression et de corrélation

Dans le but d'analyser l'effet d'un pourcentage d'hormones de croissance dans la nourriture (X) sur le poids en kg (Y) de dindons à 20 semaines, on a prélevé un échantillon aléatoire de dindons; les résultats sont indiqués dans le Tableau 2.3. Effectuez une analyse de régression et de corrélation de ces données.

TABLEAU 2.3: Données de régression de l'étude portant sur les dindons

Pourcentage d'hormones (X)	4	6	8	10	12	14	16	18	20
Poids (kg.) (Y)	4.1	4.9	4.6	5.3	5.8	5.6	6.4	6.4	6.7

Dans cet exemple, $n = 9$, $\Sigma X = 108$, $\Sigma Y = 49.8$, $\overline{X} = 12$, $\overline{Y} = 5.53$, $\Sigma X^2 = 1536$, $\Sigma Y^2 = 281.88$, $\Sigma XY = 635.2$

d'où,
$$b = \frac{635.2 - \frac{(108)\ (49.8)}{9}}{1536 - \frac{(108)^2}{9}} = 0.157 \text{ et}$$

$$a = \frac{49.8}{9} - \frac{(0.157)\ (108)}{9} = 3.649 \quad .$$

Par conséquent, l'équation de régression s'écrit:

$$\hat{Y} = 3.649 + 0.157X.$$

Pour tracer la ligne de régression estimée, trouvez la valeur de $\hat{Y}$ pour deux valeurs de X sélectionnées (2 et 20, par exemple). Quand $X = 2$, $\hat{Y} = 3.963$. Lorsque $X = 20$, $\hat{Y} = 6.789$. Joignez ces deux points à l'aide d'une règle de manière à obtenir la ligne désirée (se reporter à la Figure 2.6).

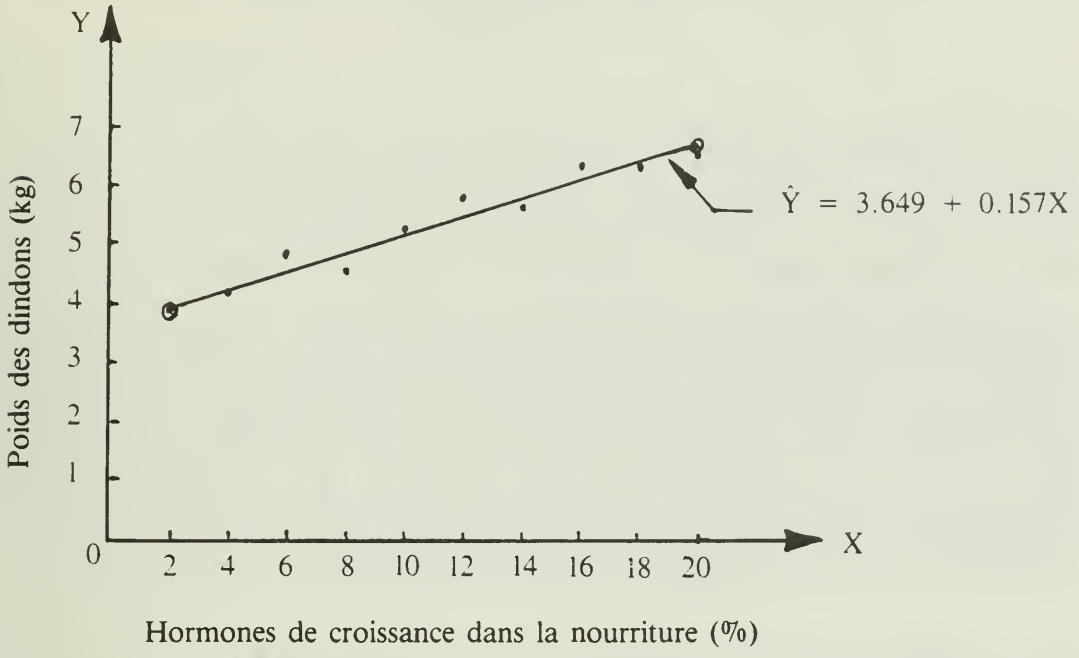


Figure 2.6 Ligne de régression pour les données du Tableau 2.3

Établissons un intervalle de prévision de 95% pour Y lorsque X = 15 (%). Quand $X = X_0 = 15$, $\hat{Y} = 6.004$,

$$s_{y.x}^2 = \frac{281.88 - (3.649)(49.8) - (0.157)(635.2)}{9 - 2} = 0.0619, \text{ et}$$

$$s_{\hat{y}}^2 = 0.0619 \left(1 + \frac{1}{9} + \frac{(15 - 12)^2}{1536 - \frac{(108)^2}{9}} \right) = 0.0711 .$$

Par conséquent, un intervalle de prévision de 95% pour Y lorsque $X = X_0 = 15$, se rend par $6.004 \pm 2.3646 \sqrt{0.0711}$ ou (5.373 , 6.635) kg, où 2.3646 correspond à la valeur de $t_{.025}$ qui figure sur le Tableau 2 de l'Appendice pour $f = n - 2 = 7$ degrés de liberté.

Le coefficient de corrélation s'écrit:

$$r = \frac{635.2 - \frac{(108)(49.8)}{9}}{\sqrt{\left(1536 - \frac{(108)^2}{9}\right)\left(281.88 - \frac{(49.8)^2}{9}\right)}} = 0.9654 .$$

Pour vérifier la signification de ce coefficient de corrélation, c.-à-d., H_0: $\rho = 0$ par rapport à H_A: $\rho > 0$, on calcule la statistique d'essai pertinente comme suit:

$$t = \frac{r}{\sqrt{\frac{1 - r^2}{n - 2}}} = \frac{0.9654}{\sqrt{\frac{1 - (0.9654)^2}{9 - 2}}} = 9.795 \ .$$

On peut lire dans le Tableau 2 de l'Appendice que la valeur de $t_{.05}$ pour $n - 2 = 9 - 2 = 7$ degrés de liberté est 1.8946. Étant donné que la valeur de t calculée (9.795) est de beaucoup supérieure à la valeur de t du tableau (1.8946), on rejette H_0 et on conclut qu'il y existe, pour un niveau de signification de 5%, une corrélation positive très significative entre le pourcentage d'hormones et le poids.

L'analyse qui précède nous permet donc de conclure qu'il y existe une relation directe entre X et Y, c.-à-d., une augmentation du pourcentage d'hormones de croissance dans la nourriture se traduit par un accroissement du poids des dindons de 20 semaines. On peut tirer la même conclusion du coefficient de détermination $r^2 = (0.9654)^2 = 0.9320$, qui révèle que 93.2% de la variation du poids des dindons de 20 semaines s'explique par l'association linéaire entre le poids et le pourcentage d'hormones de croissance présentes dans la nourriture.

CHAPITRE 3

MÉTHODES D'ÉCHANTILLONNAGE

3.1 Introduction

Le but de toute recherche, qu'il s'agisse d'une expérience en laboratoire, d'une inspection par échantillonnage de produits alimentaires ou d'une enquête, est d'estimer ou de comparer les caractéristiques d'un lot ou d'une population ou de décider d'accepter ou de rejeter un lot. Pour ce faire, nous aimerions, idéalement, pouvoir nous servir de toute la population pour étudier les caractéristiques qui nous intéressent. Malheureusement, dans la plupart des cas, un dénombrement de toute la population ou une inspection complète (100%) est soit impossible, soit irréalisable, ou encore trop coûteux, sans parler du temps considérable que cela suppose, ni de la fatigue et de l'ennui qui en résultent pour le personnel chargé de l'inspection. Par conséquent, on fait des généralisations et des inférences sur une population en observant le comportement et les propriétés d'un nombre relativement restreint d'unités prises dans cette population, que l'on appelle un échantillon.

La théorie de l'échantillonnage a pour but d'améliorer l'échantillonnage, c.-à-d., d'élaborer des méthodes de sélection de l'échantillon et d'estimation qui permettront de recueillir les meilleurs informations au plus bas coût possible. On s'attend que les estimations qui s'ensuivent soient suffisamment exactes et précises. L'exactitude représente le rapport étroit entre la valeur mesurée et la valeur réelle, et la différence absolue entre la valeur prévue et la valeur réelle définit la mesure de l'exactitude. On évalue la précision d'une estimation selon le degré de variabilité d'une mesure à l'autre et on la désigne par σ^2. Ainsi, une estimation sera exacte et sans biais si la valeur réelle et la valeur prévue sont identiques et elle sera précise si la valeur de σ^2 est petite.

3.2 Types de procédés d'échantillonnage

La nature et le but d'une enquête détermine le type de procédé d'échantillonnage à utiliser. D'une façon générale, il y a quatre facteurs à déterminer:

- **Priorité d'échantillonnage:** dans le cas d'une multitude de produits présentant des risques divergents ou des caractéristiques critiques, établir la priorité des lots à échantillonner de manière à répartir le mieux possible les ressources.
- **Fréquence d'échantillonnage:** établir le niveau d'inspection requis pour chaque produit ou activité.
- **Taille de l'échantillon:** déterminer la quantité de produits à choisir pour l'inspection pour chaque activité ou lot.

- **Méthode de sélection d'un échantillon:** déterminer la méthode statistique la plus appropriée pour choisir la taille de l'échantillon désignée.

3.2.1 Priorité d'échantillonnage

La priorité d'échantillonnage est une fonction de plusieurs facteurs, dont:

- l'importance relative des produits sur le marché domestique, ainsi que sur le marché des importations et des exportations
- les caractéristiques des produits et leurs répercussions
- l'ordonnancement et le volume de la production, ainsi que l'historique de la qualité/conformité d'un produit
- les coûts relatifs à l'inspection
- l'emplacement physique d'un produit
- la disponibilité des ressources affectées à l'inspection
- les répercussions sur le commerce
- les besoins et les attentes du consommateur

On peut établir la priorité d'échantillonnage en élaborant un modèle à plusieurs variables qui prend en compte les facteurs essentiels, ou en faisant appel à des techniques de gestion pour répartir le mieux possible les ressources financières et humaines qui nous sont disponibles.

3.2.2 Fréquence d'échantillonnage

Comme dans le cas de la priorité d'échantillonnage, l'établissement de la fréquence d'échantillonnage dépend de beaucoup de facteurs, dont:

- le volume, la fréquence et l'ordonnancement de la production
- l'historique de la qualité/conformité du produit
- le système de gestion de la qualité de l'organisme
- la distance à parcourir par l'inspecteur pour faire son travail
- la présentation du produit et le plan d'aménagement des installations
- les ressources et les coûts relatifs à l'inspection à effectuer
- les exigences en matière de réglementation et les besoins du consommateur

D'une façon générale, on commence par une inspection normale, proportionnée aux ressources disponibles, pour passer ensuite à un mode d'inspection rigoureux ou réduit selon que la qualité du produit se détériore ou s'améliore.

Une méthode probabiliste, comme celle qui suit, peut parfois être utilisée avantageusement pour établir la fréquence d'échantillonnage:

$$\text{Fréquence d'échantillonnage} = f = \log p_2 / \log p_1$$

où

p_1 = la probabilité de ne pas détecter la présence d'un 'problème' dans le premier échantillon prélevé une fois que le 'problème' s'est manifesté,

p_2 = la probabilité de ne pas détecter la présence d'un tel 'problème' dans f échantillons consécutifs.

Si, par exemple, $p_1 = 0.83$ et $p_2 = 0.05$, f est égal à 16. Cela signifie que, même si la probabilité de ne pas détecter la présence d'un 'problème' dans le premier échantillon est aussi élevée que 0.83, la probabilité que ce 'problème' passe inaperçu dans 16 échantillons consécutifs n'est pas supérieure à 0.05.

3.2.3 Taille de l'échantillon

La question qui revient le plus souvent dans le cadre d'une enquête concerne la façon de déterminer la taille de l'échantillon. La détermination de la taille de l'échantillon est nécessaire pour deux raisons: estimer les paramètres d'un lot ou d'une population et prendre la décision d'accepter ou non un lot. Nous examinerons dans le Chapitre 4 cette dernière raison, appelée échantillonnage de réception.

Une des difficultés fondamentales que pose la solution de la plupart des problèmes liés à la "taille de l'échantillon" vient le plus souvent de ce qu'on ne sait pas ce que l'on veut et qu'on ne dispose pas de tous les renseignements nécessaires pour faire les calculs. Avant de parvenir à fixer la taille d'un échantillon, on doit répondre à trois questions:

1. Dans quelle mesure l'expérience variera-t-elle?
2. Quelle différence y aura-t-il entre les estimations et les valeurs réelles? Quelles seront les différences entre les traitements?
3. Quelle exactitude attend-on des estimations?

On a imaginé différentes méthodes pour déterminer rapidement la taille d'un échantillon, comme l'extraction de la racine carrée du nombre d'éléments dans un lot ou le prélèvement d'un pourcentage fixe d'éléments, 10% par exemple, dans un lot. Bien que ces techniques soient faciles à utiliser, elles ne reposent pas sur une considération statistique de l'expérience. Nous examinerons dans la Section 3.3 une formule de base pour déterminer la taille d'un échantillon à des fins d'estimation, dans le cas d'une distribution normale.

3.2.4 Méthode de sélection d'un échantillon

Après avoir défini la taille de l'échantillon, comment choisit-on un échantillon aléatoire et représentatif? On peut se servir, à la base, d'une méthode d'échantillonnage au jugé ou au hasard ou choisir une des méthodes d'échantillonnage probabiliste décrites dans la Section 3.4.

3.3 Détermination de la taille d'un échantillon: Une formule

Voici une formule simple qui permet de déterminer la taille d'un échantillon lorsque la caractéristique à mesurer est distribuée à peu près normalement:

$$n = \frac{N}{1 + \frac{e^2 (N - 1)}{Z^2 (p) (1 - p)}}$$

où

n = la taille de l'échantillon

N = la taille du lot

Z = la valeur normale type selon un niveau de confiance donné (par ex., Z = 1.96 selon un niveau de confiance de 95%)

e = la marge d'erreur que l'analyste consent à tolérer entre l'estimation et la valeur réelle

p = la proportion estimée d'éléments défectueux .

On remarquera que la valeur de e est choisie arbitrairement par l'analyste; plus la valeur de e est petite, plus l'échantillon sera grand. On obtient la valeur de p à partir de la qualité moyenne de production globale.

Lorsque la valeur de l'écart-type σ est connue, la formule correspondante s'écrit:

$$n = \frac{N}{1 + \frac{e^2 (N - 1)}{Z^2 \sigma^2}} .$$

Dans les cas où la valeur de N est très élevée ou que la fraction de l'échantillonnage $\frac{n}{N}$ est presque égale à un, on peut ramener les formules susmentionnées à:

$$n = \frac{Z^2 (p) (1 - p)}{e^2} = \frac{Z^2 \sigma^2}{e^2} .$$

On peut constater l'effet conjugué des paramètres e et p sur la détermination de la taille de l'échantillon en consultant le Tableau 3.1. Le tableau indique la valeur respective de n pour certaines valeurs de N, p et e selon un niveau de confiance de 95%, c.-à-d., lorsque Z = 1.96.

TABLEAU 3.1: Effet des paramètres sur la taille de l'échantillon ($1 - \alpha = 0.95$)

e	.01			.10		
N \ p	.01	.10	.20	.01	.10	.20
500	217	437	463	4	33	55
1,000	276	776	861	4	34	58
2,000	320	1268	1510	4	35	60
5,000	354	2045	2758	4	35	61
10,000	367	2570	3807	4	35	62
50,000	378	3234	5474	4	35	62

Exemple 3.1 Détermination de la taille de l'échantillon

Si l'on constate sur une longue période, qu'en moyenne, 4% d'un produit manufacturé est défectueux, quelle devra être la taille minimale d'un échantillon tiré d'un lot de 400 unités de ce produit pour que l'on soit certain à 95% que la marge d'erreur de l'estimation moyenne ne dépassera pas 3%?

Dans cet exemple, $p = 0.04$, $1 - p = 0.96$, $N = 400$, $e = 0.03$ et $Z = 1.96$. Alors,

$$n = \frac{N}{1 + \dfrac{e^2 (N - 1)}{Z^2 (p) (1 - p)}} = \frac{400}{1 + \dfrac{(0.03)^2 (399)}{(1.96)^2 (0.04) (0.96)}} = 116.47 .$$

Compte tenu des conditions énoncées, il faudra donc tirer un échantillon d'au moins 117 éléments d'un lot contenant 400 unités du produit.

3.4 Méthodes de sélection d'un échantillon

Après avoir fixé la taille de l'échantillon, on peut avoir recours à des méthodes d'échantillonnage non aléatoire ou à des méthodes d'échantillonnage probabiliste pour faire la sélection proprement dite de l'échantillon. Voici certaines méthodes d'échantillonnage non probabiliste: échantillonnage au jugé, échantillonnage au hasard, échantillonnage pratique, échantillonnage par unités naturelles, échantillonnage par la méthode des quotas, etc.

Bien que certaines situations ne permettent de faire appel qu'à des méthodes d'échantillonnage non probabiliste, il faut veiller à conserver le caractère aléatoire et représentatif d'un échantillon. On compte au nombre des méthodes d'échantillonnage probabiliste: l'échantillonnage aléatoire simple, l'échantillonnage systématique, l'échantillonnage stratifié, l'échantillonnage en grappes, etc.

3.4.1 Échantillonnage aléatoire simple

L'échantillonnage aléatoire simple est celui qui est tiré à partir d'un lot ou d'une population d'une manière aléatoire et de façon que chacun des éléments du lot ait une probabilité égale et indépendante d'être choisi. Il peut se faire à l'aide de tables de nombres aléatoires. Il existe de nombreux types de tables de ce genre. L'une de ces tables est présentée dans le Tableau 4 de l'Appendice; elle est extraite d'une publication de la Rand Corporation intitulée "A Million Random Digits with 100,000 Normal Deviates" (Free Press of Glenco, New York, 1955). À titre d'exemple, nous avons pris les 150 premiers nombres aléatoires de ce Tableau 4 que nous reproduisons dans le Tableau 3.2 ci-dessous. Nous invitons ceux qui travaillent avec de grandes populations ou qui ont à répéter souvent la sélection de nombres aléatoires à consulter la table plus détaillée de la Rand Corporation.

Prélèvement d'un échantillon aléatoire simple

Supposons qu'on veuille prélever un échantillon aléatoire simple de huit boîtes au sein d'un lot de 90 boîtes. Les boîtes du lot sont numérotées de 1 à 90.

TABLEAU 3.2: Nombres aléatoires

93108	77033	68325	10160	38667	62441
87023	94372	06164	30700	28271	08589
83279	48838	60935	70541	53814	95588
05832	80235	21841	35545	11148	34775
17308	88034	97765	35959	52843	44895

En commençant en haut de la première colonne du Tableau 3.2, séparez les nombres par groupes de deux chiffres pour obtenir le numéro des boîtes suivantes à prélever:

93 10 87 70 33 68 32 51

On remarquera que la présentation des nombres par groupes de cinq chiffres ne vise qu'à en faciliter la lecture. Comme le numéro 93 ne correspond à aucune boîte du lot, on n'en tient tout simplement pas compte. Nous choisissons donc sur la première rangée le numéro aléatoire suivant pour remplacer 93, soit 01. De la même façon, si un nombre était répété, nous le remplacerions par le nombre suivant. Les huit boîtes qui sont finalement choisies pour faire partie de l'échantillon portent donc les numéros:

10 87 70 33 68 32 51 1

3.4.2 Échantillonnage aléatoire stratifié

Pour obtenir un échantillon aléatoire stratifié, on réunit les éléments de la population en groupes qui s'excluent mutuellement, appelés les strates; on choisit ensuite un échantillon aléatoire simple dans chaque strate. Il y a trois raisons principales

pour lesquelles l'échantillonnage aléatoire stratifié a souvent une meilleure valeur informative, compte tenu d'un coût donné:

1. Les données de chaque strate sont plus homogènes que celles de l'ensemble de la population.
2. Le travail administratif étant allégé, l'échantillonnage aléatoire stratifié est généralement moins coûteux que l'échantillonnage aléatoire simple.
3. Quand on a recours à l'échantillonnage stratifié, on peut obtenir des estimations distinctes de certains paramètres de la population pour chaque strate, sans échantillonnage additionnel.

Comme les éléments de chaque strate sont plus homogènes, les estimateurs tirés de l'échantillonnage stratifié ont une variance moins élevée que les estimateurs d'un échantillon aléatoire simple de même taille.

Dans un poste d'emballage d'oeufs, par exemple, les boîtes d'oeufs sont classées par palettes en fonction de la catégorie des oeufs; la population est donc divisée en strates (c.-à-d., petit, moyen, gros, extra gros) et l'inspection peut se faire par échantillonnage aléatoire stratifié.

3.4.3 Échantillonnage systématique

L'échantillon qu'on obtient en choisissant au hasard un élément des premiers k éléments d'une population et chaque k^e élément par la suite s'appelle échantillon systématique un sur k. Supposons qu'on ait N éléments d'une population numérotés dans l'ordre de 1 à N et qu'on veuille en tirer un échantillon de taille n. On détermine un nombre entier k, appelé intervalle d'échantillonnage, qui est le nombre entier le plus proche de N/n, c.-à-d., $k \simeq \frac{N}{n}$, puis on choisit au hasard un nombre c entre 1 et k inclusivement. L'échantillon aléatoire systématique requis comprend donc les éléments numérotés

$$c,\ c + k,\ c + 2k,\ \ldots,\ c + (n - 1)k\ .$$

L'échantillonnage systématique peut avantageusement remplacer l'échantillonnage aléatoire simple parce qu'il est plus facile à prélever, qu'il est moins sujet aux erreurs et qu'il a une plus grande valeur informative pour un coût donné.

Un agriculteur, par exemple, pourra utiliser un échantillon systématique un sur 10 pour déterminer la qualité de l'eau d'érable de ses érables à sucre, si le nombre total d'arbres N lui est inconnu et qu'il ne peut donc prélever un échantillon aléatoire simple.

3.4.4 Échantillonnage en grappes

Un échantillon en grappes est un échantillon aléatoire simple dans lequel chaque unité d'échantillonnage prend la forme d'un groupe d'éléments, la grappe. La

population est divisée en grappes de manière que les grappes se ressemblent le plus possible. L'hétérogénéité de la population se reflète dans chaque grappe.

L'échantillonnage aléatoire en grappes est moins coûteux que l'échantillonnage aléatoire simple ou stratifié si la création de la base ou la liste de toutes les unités de la population occasionne des dépenses très élevées ou si les coûts relatifs au prélèvement des observations augmentent suivant la distance qui sépare les unités.

La première étape d'un échantillonnage en grappes consiste à préciser approximativement les grappes. Comme les éléments d'une grappe sont souvent près les uns des autres, ils ont tendance à avoir des caractéristiques similaires. Ainsi, la multiplication de nouvelles mesures au sein d'une grappe ne fournira pas nécessairement beaucoup plus d'information concernant un paramètre quelconque de la population. D'une façon générale, le nombre d'éléments dans une grappe sera peu élevé par rapport à la taille de la population, alors que le nombre de grappes dans l'échantillon sera assez élevé.

Supposons, par exemple, que l'Office de commercialisation du dindon veuille estimer le volume annuel de dindons achetés par ménage dans un comté à population faible disséminée sur tout le territoire. Comme les déplacements d'un ménage à l'autre occasionnent des dépenses élevées, on dresse une liste des 15,000 ménages du comté en les groupant en 600 grappes géographiques semblables, de 25 ménages chacune, et on y prélève un échantillon aléatoire simple de 30 grappes.

3.5 Échantillonnage d'ensemble

L'échantillonnage d'ensemble désigne l'échantillonnage de produits en vrac. Les produits peuvent se présenter sous forme gazeuse, liquide ou solide. Ils peuvent être homogènes (non formés d'éléments distincts), comme l'acide d'un contenant, ou formés d'éléments distincts, ce qui est généralement le cas des produits qu'on trouve en vrac à l'état naturel, comme les solides et les liquides qui sont expédiés par grands réservoirs, wagons et bateaux ou qui sont stockés. Les produits peuvent se présenter sous forme de piles sans subdivisions assimilables à des unités d'échantillonnage. Ils peuvent également être emballés, mis en sac ou subdivisés en unités d'échantillonnage distinctes qui se prêtent bien à une activité d'échantillonnage courante. Enfin, les produits peuvent se trouver dans une **situation statique** ou dans une **situation dynamique.**

Les situations statiques comprennent les produits entassés en vrac dans l'entrepôt d'une usine ou sur des docks, des produits transportés en vrac par barge, rail ou route, des produits entassés ou ensilés dans des magasins ou dans des exploitations agricoles, etc. Du strict point de vue de la théorie de l'échantillonnage, il est impossible de tirer un échantillon représentatif d'un tas statique car on enfreint nécessairement l'une des règles essentielles de l'échantillonnage qui stipule que chaque parcelle doit avoir une chance égale d'être choisie; à moins que le tas ne puisse passer par un dispositif d'échantillonnage, prendre une forme conique ou être complètement sectionné, cela est évidemment impossible. Les parcelles qui se trouvent au centre ou à la base du tas n'ont peut-être aucune chance d'être choisies. Par ailleurs, les problèmes de ségrégation des tas statiques

sont bien connus, et cette ségrégation même peut avoir une incidence sur la distribution de caractéristiques comme la composition chimique, les propriétés physiques, etc. La ségrégation peut être le résultat de variations de la taille des éléments ou de la densité des composants d'un mélange.

Les situations dynamiques s'observent en usine (emplissage d'une aire de stockage), à un point de chargement d'un vraquier, à un point de livraison d'un cargo transportant des produits en vrac, etc. Dans ces situations, on peut prélever des échantillons sur les convoyeurs à courroie par des moyens mécaniques ou autres, ou l'échantillonnage peut parfois se faire à un point de chute du produit. Les situations dynamiques posent moins de problèmes, et on peut normalement appliquer à celles-ci des plans d'échantillonnage existants.

L'échantillonnage d'ensemble peut avoir un des objectifs suivants ou plus:

- la caractérisation des produits pour ce qui est de la quantité, du contenu, de la valeur, de la catégorie, de l'homogénéité, etc.
- l'estimation de la valeur moyenne des caractéristiques du produit ainsi que de leur variabilité
- l'acceptation de chaque lot
- le contrôle pendant le traitement
- la conformité aux spécifications et aux tolérances
- l'établissement de procédures d'échantillonnage uniformes

3.5.1 Échantillonnage de produits ayant subi une ségrégation

Habituellement, pour échantillonner des produits en vrac, on prélève par intermittence plusieurs échantillons du lot, on les mélange en un échantillon composite unique et, au besoin, on donne à l'échantillon une taille convenable au travail en laboratoire. Si les produits se présentent par lots de contenants et qu'on connaît leur mode de ségrégation, on a souvent recours à un plan d'échantillonnage à emboîtements dans lequel on choisit successivement et d'une manière aléatoire un certain nombre de contenants au sein de ces lots et d'échantillons au sein de ces contenants. Si le produit a tendance à se stratifier, on peut prélever des échantillons aléatoires dans chaque strate ou au sein d'un certain nombre de strates choisies aléatoirement.

Par exemple, dans l'échantillonnage à deux degrés d'une population en vrac comprenant N éléments primaires, chaque élément étant composé de M échantillons possibles, le plan d'échantillonnage suppose qu'on choisisse au hasard n éléments primaires et qu'on prélève dans chacun m échantillons de façon aléatoire. L'échantillon composite ainsi obtenu est ensuite ramené à des proportions adaptées au travail en laboratoire.

Bien qu'on ait fait peu de travaux statistiques sur la normalisation de l'échantillonnage des produits en vrac, certaines situations ont été bien étudiées. C'est le cas de l'échantillonnage des engrais. L'*Association of Official Agricultural*

Chemists donne dans *The Official Methods of Analysis* les directives suivantes pour le prélèvement d'échantillons d'engrais et la formation d'un échantillon composite:

> Se servir d'un tube rainuré simple ou double, ou d'un tube rainuré et d'une tige, à extrémité conique. Prélever des échantillons comme suit: déposer les sacs horizontalement et prélever en diagonale des carottes d'une extrémité à l'autre. Si le lot comporte 10 sacs ou plus, tirer un échantillon par groupe de 10 sacs. Si le lot comporte moins de 10 sacs, prélever au moins une carotte dans chaque sac. Si l'engrais se présente en vrac, prélever au moins 10 carottes en divers points. L'échantillonnage des livraisons en vrac peut se faire pendant le chargement ou le déchargement par prélèvements au point de chute de la courroie ou de la goulotte. Si les paquets sont de petite taille (10 livres ou moins), prendre un paquet au complet. Ramener l'échantillon composite à la quantité requise, de préférence avec un échantillonneur à riffles ou par mélange (sur une toile cirée) et quartage. Mettre l'échantillon dans un contenant hermétique.

Les instructions sur la préparation de l'échantillon en vue d'une analyse en laboratoire se lisent comme suit:

> Ramener l'échantillon brut à une quantité suffisante pour fins d'analyse ou broyer au moins 0.5 lb d'échantillon non tamisé. S'il s'agit d'engrais humides, broyer l'échantillon de façon qu'il puisse passer par un tamis à mailles circulaires de 1 mm ou par un tamis n° 29. S'il s'agit de mélanges secs sensibles à la ségrégation, faire le broyage à l'aide d'un tamis n° 40. Broyer le plus rapidement possible afin d'éviter toute diminution ou augmentation de la teneur en humidité. Mélanger à fond et conserver dans des flacons bouchés avec soin.

3.6 Plan général d'une méthode d'échantillonnage

Il faut concevoir et planifier judicieusement le processus d'échantillonnage. Un bon plan type englobe tous les éléments requis et prévoit des dispositions pour assurer l'uniformité de leur application. Nous vous proposons ci-après le plan général d'une méthode d'échantillonnage applicable aux produits agricoles et alimentaires, qui s'inspire de la norme internationale ISO/7002. On peut modifier ce plan en fonction des exigences particulières d'une situation.

Voici une liste générale des éléments requis/recommandés pour établir le plan d'une méthode d'échantillonnage type:

- Objectif(s)
- Principe de la méthode d'échantillonnage: étapes essentielles de la méthode à utiliser, nature du produit à échantillonner, but de l'échantillonnage, plan d'échantillonnage approprié
- Arrangements administratifs: personnel d'échantillonnage, représentation des intéressés, mesures de précaution par rapport à la sécurité, à la santé et à la sûreté, apposer une signature au rapport d'échantillonnage

- Identification et inspection générale du lot avant l'échantillonnage: identification du lot avant l'échantillonnage, conditions et caractéristiques du lot et de son environnement, séparation du lot en unités homogènes, méthode de marquage des unités du lot en vue du prélèvement d'un échantillon aléatoire, méthodes d'acceptation/de rejet d'un lot
- Matériel d'échantillonnage et conditions ambiantes
- Contenants de l'échantillon et emballage spécial: propreté, qualité, convenance, solidité
- Procédures d'échantillonnage: taille de l'échantillon, méthode d'échantillonnage par accroissement, préparation d'un échantillon d'ensemble, d'un échantillon composite et d'un échantillon réduit, sélection d'un échantillon de produits préemballés
- Emballage, scellage et marquage des échantillons et des contenants d'échantillons
- Mesures de précaution pendant l'entreposage et le transport des échantillons
- Rapport d'échantillonnage: détails administratifs, détails relatifs aux unités emballées ou aux contenants du lot, produit échantillonné, méthode d'échantillonnage, préparation et scellage des échantillons
- Pièces jointes, au besoin: modèles de rapport, notes de mise en garde, renvois aux dispositions légales

CHAPITRE 4

ÉCHANTILLONNAGE DE RÉCEPTION

4.1 Introduction

L'échantillonnage de réception désigne l'opération par laquelle on accepte ou on rejette un lot à la suite de l'inspection d'un échantillon choisi selon un plan d'échantillonnage préétabli. Le **plan d'échantillonnage** précise le nombre d'unités à échantillonner, les critères d'acceptation et de rejet ainsi que les probabilités et les risques liés à la réception. Le but de l'échantillonnage de réception n'est pas de donner une estimation de la qualité du lot, mais de prendre une décision quant à la destination du lot. Les plans d'échantillonnage de réception ne permettent pas d'exercer un contrôle direct sur la qualité; ils sont essentiellement des outils de vérification qu'on utilise pour faire en sorte que le résultat d'un processus soit conforme aux exigences.

Les plans d'échantillonnage reposent sur plusieurs caractéristiques liées à la qualité et le choix d'un plan ou d'un autre dépend surtout de la nature du produit et de l'objet de l'inspection. Le choix d'un plan d'échantillonnage adéquat et approprié est une tâche très importante mais souvent difficile car il met en jeu de nombreux facteurs différents, y inclus sa facilité d'utilisation, la protection assurée, les travaux d'inspection requis, les coûts relatifs à l'inspection et la capacité du plan de distinguer un bon lot d'un mauvais.

4.2 Classification des plans d'échantillonnage

Il y a différentes façons de classer les plans d'échantillonnage de réception. L'une des plus importantes se fonde sur les attributs et les variables. Les **variables** sont, naturellement, des caractéristiques de qualité qui se mesurent à l'aide d'une échelle numérique continue. Les **attributs** sont des caractéristiques de qualité qui sont exprimées suivant un principe dichotomique: "défectueux, non défectueux", "correct, incorrect". La classification suivante des plans d'échantillonnage porte sur les indices et les risques de qualité. Certains indices de qualité importants se définissent comme suit:

- **Niveau de qualité acceptable (NQA)**
  C'est le niveau de qualité qui, aux fins de l'inspection par échantillonnage, correspond à la limite d'une qualité moyenne de production satisfaisante. La **qualité moyenne de production** représente le pourcentage de qualité moyenne pour une période ou un volume de production donné. Le NQA se rapporte au **risque α (alpha)**, appelé aussi **"risque à la production"**, soit la probabilité de commettre une erreur de type I, ou de rejeter un bon lot.

- **Qualité limite (QL)**
  C'est le niveau de qualité qui, aux fins de l'inspection par échantillonnage, correspond à la limite d'une qualité moyenne de production non acceptable.

On l'appelle également le "**pourcentage de produits défectueux toléré dans un lot (LTPD)**". La QL se rapporte au **risque ß (bêta)**, appelé aussi "**risque à la consommation**", soit la probabilité de commettre une erreur de type II, ou d'accepter un mauvais lot.

- **Qualité moyenne de sortie (QMS)**
  C'est le niveau de qualité moyen prévu du produit sortant compte tenu de la valeur correspondant à la qualité du produit entrant, calculé pour tous les lots acceptés et tous les lots rejetés après que ceux-ci ont été soumis à une inspection complète et que tous les éléments non-conformes ont été remplacés par de bons éléments. La "**limite de qualité moyenne de sortie (LQMS)**" se définit comme la QMS maximale pour toutes les valeurs possibles de qualité du produit entrant, compte tenu d'un plan d'échantillonnage de réception donné et de certaines spécifications relatives au rejet des lots.

4.3 Caractérisation d'un plan d'échantillonnage

Chaque plan d'échantillonnage de réception est caractérisé par les facteurs suivants: la taille de l'échantillon (n), le critère d'acceptation (Ac) et le critère de rejet (Re), et la probabilité de réception (P_a). La **taille de l'échantillon** correspond au nombre d'éléments soumis à l'inspection. Le **critère d'acceptation** équivaut au nombre maximal d'éléments défectueux (ou de défauts) dans l'échantillon en deçà duquel le lot sera accepté. Le **critère de rejet** équivaut au plus petit nombre d'éléments défectueux (ou de défauts) dans l'échantillon au-delà duquel le lot sera rejeté. La **probabilité de réception** d'un plan d'échantillonnage est égale au pourcentage d'échantillons acceptés, parmi une longue série d'échantillons, qui permet la réception du produit. En représentant graphiquement la probabilité de réception pour tous les niveaux possibles de pourcentage d'éléments défectueux, on obtient la **courbe caractéristique d'efficacité (CE).** Grâce à la courbe CE d'un plan d'échantillonnage, on peut quantifier les risques, les exprimer numériquement et établir le nombre de produits qui devraient être acceptés si les normes de qualité sont respectées et rejetés si les normes ne le sont pas.

En pratique, on se sert d'une probabilité de réception de 95% et de 10% respectivement pour un certain NQA et une QL donnée. Ainsi, la probabilité de réception d'un lot de NQA est de $1 - \alpha = 0.95$, et de $ß = 0.10$ pour un lot de QL. La Figure 4.1 représente la courbe CE de ces valeurs de la probabilité de réception (P_a), correspondant à un NQA de 2% et à une QL de 8%, selon un échantillon de taille fixe $n = 100$ et un critère d'acceptation $Ac = 4$. On peut constater que, dans ces circonstances, le produit sera accepté dans 95% des cas si le taux d'éléments défectueux est de 2%, et dans seulement 10% des cas, si le taux est de 8%.

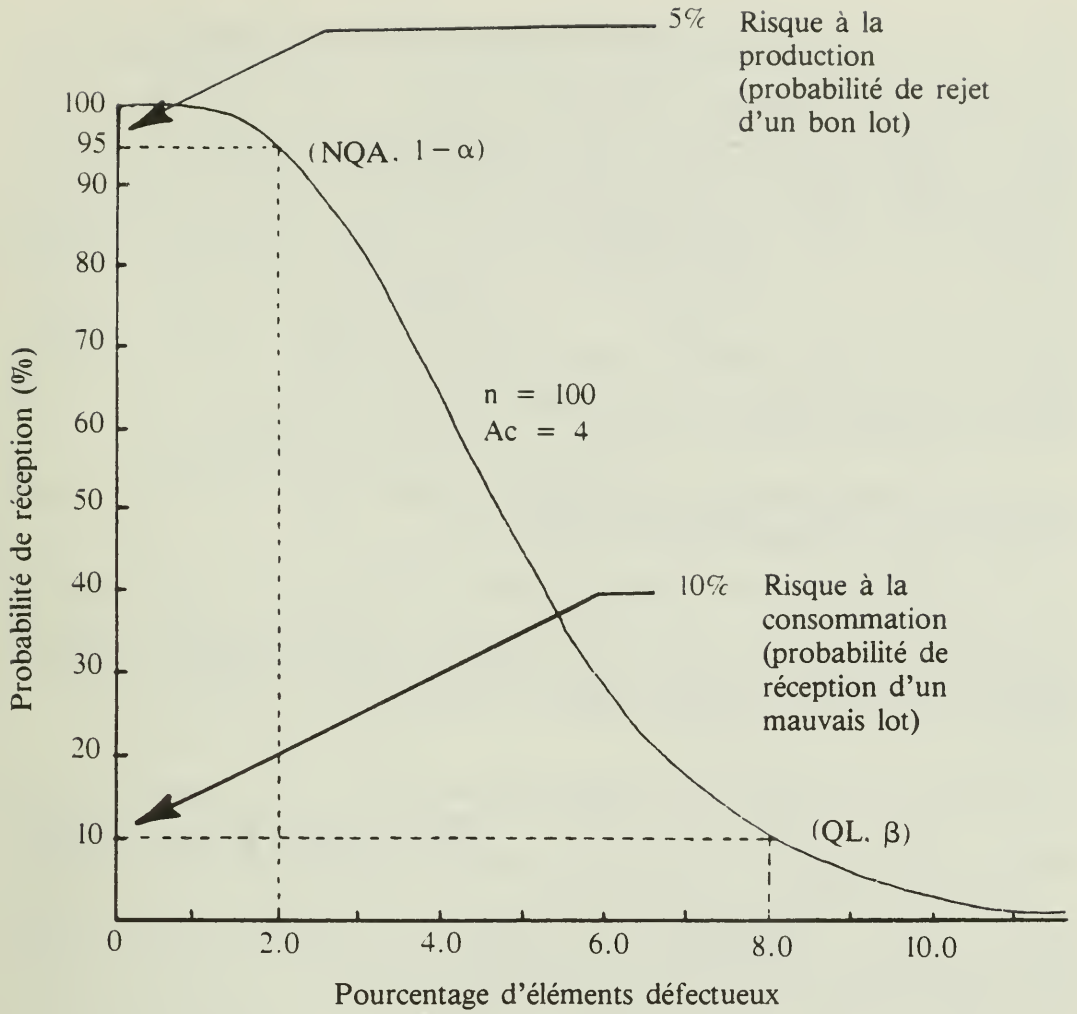


Figure 4.1 Courbe CE d'un plan d'échantillonnage simple

4.4 Choix d'un plan d'échantillonnage

Il y a beaucoup de facteurs à considérer dans le choix d'un plan d'échantillonnage qui convient à une situation particulière: (i) l'objectif de l'inspection, (ii) la nature du produit, (iii) le type de méthodes d'essai en vigueur, et (iv) la nature des lots à échantillonner en sont des exemples. L'inspection peut avoir pour objet de décider d'accepter ou de rejeter un lot, de mesurer la qualité moyenne d'un lot ou de déterminer la variabilité ou l'uniformité d'un produit. Des facteurs liés aux matériaux, tels leur homogénéité, la taille de l'unité, la constance en répondant aux spécifications établies et le coût, peuvent influer sur la taille de l'échantillon. La méthode d'essai peut être déterminante dans le choix de la procédure d'échantillonnage; elle peut, par exemple, prendre en compte les défauts critiques plutôt que les défauts secondaires, elle peut ou non comporter la destruction du produit, ou elle peut encore faire appel à un investissement de ressources considérable. La taille du lot peut avoir un effet sur le type de distribution servant à établir un plan d'échantillonnage, alors que la composition du lot et la distribution plus ou moins aléatoire de chaque unité dans le lot peuvent influer sur le type de plan d'échantillonnage à utiliser.

Après avoir pris en considération les facteurs préalables à l'échantillonnage, on doit choisir entre un échantillonnage par attributs et un échantillonnage par variables. Dans l'échantillonnage de réception, on évalue plus souvent les attributs car les observations ainsi obtenues sont habituellement plus simples, moins coûteuses et exigent moins de temps que celles obtenues au moyen d'une variable. Par contre, les variables fournissent plus de renseignements sur les caractéristiques de qualité d'un produit et les échantillons peuvent alors être plus petits. L'échantillonnage par variables donne aussi des renseignements sur le degré de non-conformité et permet de désigner les secteurs où la qualité doit être améliorée. Quand les deux types d'échantillonnage peuvent être utilisés et qu'il faut en faire un choix, on retiendra le plan qui présente le plus d'avantages en termes de coût de l'inspection, de taille de l'échantillon, de facilité d'utilisation, etc.

Les plans d'échantillonnage sont généralement indexés selon un indice de qualité ou plus, soit, le NQA, le QL et la QMS.

Les plans NQA sont essentiellement conçus en fonction du producteur. Ils permettent à l'organisme de réglementation ou à l'acheteur de veiller à ce que le producteur respecte les normes de qualité et de sûreté approuvées. Un lot a de fortes chances d'être accepté pendant l'inspection par échantillonnage si le produit a un NQA ou supérieur.

Par ailleurs, les plans QL sont conçus en fonction du consommateur. On s'en sert pour veiller à ce que la probabilité de réception d'un lot de QL ou inférieure soit très faible. Ces plans sont surtout destinés à l'inspection de lots isolés.

Les plans QMS ne s'appliquent qu'aux programmes qui soumettent les lots refusés à une inspection complète (100%). Ces plans sont conçus de manière à réduire au minimum le taux d'éléments défectueux afin de réaliser une estimation pratique de la qualité moyenne de production, et ainsi diminuer autant que possible la quantité de travail liée à l'inspection totale moyenne pour un NQMS donné. À moins d'avoir recours à un contrôle rectificatif, le concept de NQMS perd tout son sens.

On peut, au besoin, établir un plan d'échantillonnage qui repose à la fois sur le NQA et la QL. Un plan peut aussi s'appuyer uniquement sur le NQA tout en nous permettant de prévoir facilement la QL, ou vice versa. On peut se servir de formules mathématiques de base pour obtenir le plan requis ou choisir celui-ci dans les tables, les documents ou les normes d'échantillonnage statistiques existants. Par souci de simplicité, nous limiterons notre analyse à la sélection de plans d'échantillonnage de réception à l'aide d'une des normes internationales. Étant donné que la forme la plus répandue d'inspection des produits alimentaires est le contrôle lot par lot d'un produit présenté par le producteur à l'organisme de réglementation, nous décrirons dans la Section 4.6 la méthode de sélection d'un plan d'échantillonnage par attributs indexé selon le NQA qui se fonde sur la norme internationale ISO/2859-1.

4.5 Systèmes de plans d'échantillonnage tout faits

Nous avons dressé dans cette section, à l'intention des lecteurs plus ambitieux, une liste de documents importants qui portent sur l'échantillonnage de réception. La plupart de ces documents correspondent à des normes internationales définies par l'Organisation internationale de normalisation (ISO). Pour connaître d'autres titres, le lecteur se reportera à la bibliographie.

- **Plans indexés d'après le NQA:**
  - — ISO/2859-1: Règles d'échantillonnage pour les contrôles par attributs — Partie 1: Plans d'échantillonnage pour les contrôles lot par lot, indexés d'après le NQA. Ce document est identique à la norme MIL-STD-105D du ministère américain de la Défense.
  - — ISO/3951: Règles et tables d'échantillonnage pour les contrôles par mesures des pourcentages de défectueux. Ce document est identique à la norme MIL-STD-414 du ministère américain de la Défense.
- **Plans indexés d'après la QL:**
  - — ISO/2859-2: Règles d'échantillonnage pour les contrôles par attributs — Partie 2: Plans d'échantillonnage pour les contrôles de lots isolés, indexés d'après la QL.
  - — Tableaux de contrôle par échantillonnage de Dodge et de Romig.
- **Plans indexés d'après la QMS:**
  - — Tableaux de contrôle par échantillonnage de Dodge et de Romig.
- **Autres types de systèmes de plans d'échantillonnage:**
  - — ISO/8550: Guide pour la sélection d'un plan d'échantillonnage pour acceptation.
  - — ISO/2859-0: Règles d'échantillonnage pour les contrôles par attributs — Partie 0: Introduction générale.
  - — ISO/2859-3: Règles d'échantillonnage pour les contrôles par attributs — Partie 3: Plan d'échantillonnage successif partiel.
  - — ISO/8422: Plans d'échantillonnage progressif pour le contrôle par attributs.
  - — ISO/8423: Plans d'échantillonnage progressif pour le contrôle par mesurage de la proportion d'individus non-conformes (écart-type connu).
  - — Plans d'échantillonnage continu.
  - — Plans d'échantillonnage en chaîne.
  - — ISO/3534: Statistique — Vocabulaire et symboles
    - — Partie 1: Probabilité et termes statistiques généraux
    - — Partie 2: Maîtrise statistique de la qualité
    - — Partie 3: Plans d'expérience

4.6 Plans d'échantillonnage par attributs d'après le NQA: ISO/2859-1

La norme de l'échantillonnage de réception **ISO/2859-1** a été élaborée par le comité technique TC69 de l'Organisation internationale de normalisation, et se fonde

sur la norme **MIL-STD-105D: Règles et tableaux d'échantillonnage pour les contrôles par attributs** du ministère américain de la Défense.

La norme ISO/2859-1 propose des plans d'échantillonnage pour les contrôles par attributs fondés sur le NQA et conçus pour être appliqués à de grands lots de fabrication comportant beaucoup d'unités de produit. On y trouve également des données à l'aide desquelles on peut facilement établir le niveau de protection qui correspond à chaque plan en fonction de la QL et de la LQMS.

La norme sert avant tout à assurer le maintien de la qualité à un "niveau de qualité acceptable" ou à un niveau supérieur. Elle a été conçue de façon que, si la production respecte régulièrement le NQA, la grande majorité des lots devraient être acceptés. Le NQA correspond donc à un seuil minimal de qualité que le producteur aura avantage à atteindre et même à dépasser.

La norme présente trois types de plan d'échantillonnage: **l'échantillonnage simple, double** et **multiple.** Le choix du type de plan dépend de plusieurs facteurs, tels les antécédents de qualité, les exigences en matière de qualité, le niveau d'inspection, la taille du lot, le mode d'échantillonnage, le NQA et d'autres considérations économiques.

Échantillonnage simple:

Dans un plan d'échantillonnage simple, on prélève au hasard un seul échantillon de n éléments dans un lot. La décision relative à la réception du lot s'appuie sur les résultats obtenus pour cet échantillon. Si le nombre d'éléments défectueux rencontrés au sein de l'échantillon est inférieur ou égal au critère d'acceptation Ac, le lot est accepté. Par contre, le lot est refusé si le nombre d'éléments défectueux est égal ou supérieur au critère de rejet Re.

Échantillonnage double:

Dans un plan d'échantillonnage double, on prélève au hasard un premier échantillon de n_1 éléments et on l'inspecte. Si le nombre d'éléments défectueux est inférieur ou égal au premier critère d'acceptation Ac_1, le lot est accepté. Si le nombre d'éléments défectueux est égal ou supérieur au premier critère de rejet Re_1, le lot est refusé. Si le nombre d'éléments défectueux est supérieur à Ac_1 mais inférieur à Re_1 et qu'on ne parvient pas alors à prendre une décision, on prélève au hasard un second échantillon de n_2 éléments et on l'inspecte. Si le nombre cumulatif d'éléments défectueux du premier et du second échantillon est inférieur ou égal au second critère d'acceptation Ac_2, le lot est accepté. Par contre, celui-ci est refusé si le nombre cumulatif d'éléments défectueux est égal ou supérieur au second critère de rejet Re_2.

Le nombre moyen d'éléments inspectés dans l'échantillonnage double est généralement inférieur à celui de l'échantillonnage simple. Malgré cela, l'échantillonnage double est moins populaire que l'échantillonnage simple car il comporte plus de données à rapporter. Certains inspecteurs estiment à tort que l'échantillonnage double équivaut à donner une deuxième chance au produit. On se sert habituellement de l'échantillonnage double lorsqu'on sait que la qualité d'un lot est très bonne ou très mauvaise.

Échantillonnage multiple:

L'échantillonnage multiple s'apparente à l'échantillonnage double, à cette différence près qu'on peut avoir à inspecter plus de deux échantillons consécutifs avant de décider d'accepter ou de rejeter le lot. Le nombre d'étapes requis pour prendre une décision ferme dépend du nombre cumulatif d'éléments défectueux trouvés dans les échantillons successifs. Chaque étape comporte un critère d'acceptation et de rejet; le lot est accepté si le nombre cumulatif d'éléments défectueux est inférieur ou égal au critère d'acceptation et il est refusé si le nombre cumulatif d'éléments défectueux est égal ou supérieur au critère de rejet. Si le nombre cumulatif se situe entre les critères d'acceptation et de rejet, on prélève un autre échantillon. L'échantillonnage multiple prend fin quand un nombre préétabli d'étapes a été atteint; on fixe ce nombre en choisissant pour la dernière étape des critères d'acceptation et de rejet consécutifs; une décision s'impose alors d'elle-même. À cette étape, l'échantillon cumulatif a une taille supérieure à celle de l'échantillon équivalent des deux méthodes précédentes.

Niveaux d'inspection:

Les sept niveaux d'inspection de la norme, S-1, S-2, S-3, S-4, I, II et III, correspondent à divers degrés de discrimination; chacun d'eux fournit des échantillons de tailles différentes pour un lot de taille donnée. Dans l'ordre présenté ci-dessus, la taille de l'échantillon (et, de ce fait, sa faculté de discrimination) passe d'un minimum au niveau spécial S-1 à un maximum au niveau général III. Les niveaux S-1 à S-4 sont spéciaux en ce sens que leur utilisation se limite aux cas où il est important de ne prélever que de petits échantillons, comme les tests qui supposent la destruction de produits coûteux. Le niveau général II peut être assimilé au niveau d'inspection normal; à moins d'indication contraire, c'est ce niveau qu'on utilise au début de tout travail d'inspection.

La norme prévoit également trois niveaux de sévérité: l'inspection peut être **normale, rigoureuse** ou **réduite.** Au départ, on a recours à l'inspection normale. Si la qualité est insuffisante, on demande à l'inspecteur d'être plus sévère et de faire une inspection rigoureuse. À l'inverse, si le niveau de qualité est toujours satisfaisant, il est préférable de procéder à une inspection réduite. La norme prévoit aussi des règles de passage d'un mode d'inspection à un autre.

Taille de l'échantillon:

Un code alphabétique sert à désigner la taille de l'échantillon. Le système de lettres codes est présenté dans le Tableau 1 de la norme ISO/2859-1 ainsi que dans le Tableau 5A de l'Appendice. La lettre attribuée à un échantillon donné dépend du niveau d'inspection et de la taille du lot. Dans le Tableau 5A, les lots sont groupés verticalement par groupes de taille par rapport aux niveaux d'inspection qui sont indiqués horizontalement, de sorte qu'un lot puisse être associé à un de ces groupes et correspondre à une lettre désignant la taille de l'échantillon requis.

NQA:

Le NQA est la valeur limite qui est choisie pour séparer ce qui sera considéré comme une production de qualité moyenne de ce qui ne le sera pas; il est, de

ce fait, un indicateur de la qualité à atteindre dans la production. Un NQA réaliste doit représenter un compromis entre le savoir-faire technique du producteur, les attentes du consommateur et la qualité moyenne de production accessible. La norme ISO/2859-1 indique plusieurs valeurs pour le NQA, depuis un minimum de 0.010 jusqu'à un maximum de 1000. Les NQA de 0.010 à 10 inclusivement peuvent être exprimés en pourcentage d'éléments non-conformes (défectueux) ou en nombre de non-conformités (défauts) par centaine d'unités.Les NQA supérieurs à 10 sont exprimés uniquement en nombre de défauts par centaine d'unités. Les tableaux d'échantillonnage renferment ce qu'il est convenu d'appeler "les valeurs préférentielles du NQA". Afin de simplifier la tâche administrative, il est souhaitable d'utiliser le plus possible ces valeurs. Toutefois, si la valeur du NQA spécifiée ne correspond pas à une valeur préférentielle, on ne peut pas avoir recours aux tableaux et il faut alors établir un plan d'échantillonnage spécialement conçu pour cette valeur.

4.6.1 Présentation de plans d'échantillonnage de la norme ISO/2859-1

Voici la présentation générale des plans d'échantillonnage contenus dans la norme ISO/2859-1:

- **Tableau 1:** taille des lots en groupes; niveaux d'inspection (niveaux spéciaux: S-1, S-2, S-3, S-4 et niveaux généraux: I, II, III); lettres codes des échantillons.
- **Tableaux 2-A,B,C; 3-A,B,C; 4-A,B,C:** valeurs préférentielles du NQA; taille des échantillons pour certaines lettres codes; critères d'acceptation et de rejet; type et mode d'inspection comme suit:

Mode / Type	Normal	Rigoureux	Réduit
Simple	2-A	2-B	2-C
Double	3-A	3-B	3-C
Multiple	4-A	4-B	4-C

- **Tableaux 5-A,B:** valeurs de la LQMS pour l'inspection normale et rigoureuse. Ces tableaux indiquent la limite de qualité moyenne de sortie prévisible lorsqu'on utilise un plan NQA spécifié.
- **Tableaux 6-A,B et 7-A,B:** valeurs de la QL des plans d'échantillonnage simple pour l'inspection normale. Lorsque le plan NQA est spécifié, ces tableaux permettent de déterminer la valeur limite d'une qualité moyenne de production peu satisfaisante pour laquelle la probabilité de réception est

faible. Les valeurs de la QL sont toujours supérieures à celles du NQA, et dans certains cas de beaucoup supérieures; toutefois, l'écart entre les valeurs de la QL et du NQA se rétrécit au fur et à mesure que la taille de l'échantillon augmente.

- **Tableaux 8 et 9:** nombres limites pour l'inspection réduite et courbes de la taille moyenne des échantillons pour l'échantillonnage double et multiple respectivement.
- **Tableaux 10-A à 10-S:** courbes CE et probabilités de réception exactes pour chaque lettre code des échantillons.

Nous avons reproduit dans l'Appendice quelques spécimens de pages de la norme pour exposer plus clairement l'application du document. Les Tableaux 1, 2-A et 3-A de la norme correspondent respectivement aux Tableaux 5A, 5B et 5C de l'Appendice.

4.6.2 Marche à suivre pour choisir un plan d'échantillonnage dans la norme ISO/2859-1

Voici toutes les étapes ordonnées à suivre pour se servir de la norme ISO/2859-1:

1. Fixer la taille du lot (N) qui doit être échantillonné et inspecté. Il n'est pas nécessaire que l'échantillon corresponde à un lot de fabrication complet.
2. Choisir le niveau d'inspection. En général, on recommande d'utiliser au début le niveau II.
3. À l'aide des informations des étapes 1 et 2, se reporter au Tableau 1 de la norme ISO/2859-1 (Tableau 5A de l'Appendice) pour trouver la lettre code de l'échantillon correspondante, c.-à-d., A, B, . . ., ou R, par ordre croissant de taille.
4. Choisir le type d'échantillonnage: simple, double ou multiple.
5. Décider si l'on doit commencer par une inspection normale (majorité des cas), rigoureuse ou réduite.
6. Choisir la base de l'inspection, c.-à-d., le pourcentage ou nombre d'éléments défectueux ou le pourcentage ou nombre de défauts.
7. Fixer, dans la mesure du possible, le NQA à partir des valeurs préférentielles du NQA indiquées dans le document. Les valeurs inférieures ou égales à 10.0% peuvent être exprimées en pourcentage d'éléments défectueux ou en nombre d'éléments défectueux par centaine d'unités; les valeurs supérieures à 10.0% ne sont exprimées qu'en nombre d'éléments défectueux par centaine d'unités.
8. Pour le mode et le type d'échantillonnage, ainsi que le NQA et la lettre code de l'échantillon choisis dans les étapes 5, 4, 7 et 3 respectivement, se reporter au tableau approprié dans le document. Les critères d'acceptation et de rejet sont généralement indiqués dans le corps du tableau à droite des différentes tailles d'échantillon (n).

9. On peut alors remarquer un point ou une flèche. Un point signifie qu'il faut utiliser un plan d'échantillonnage simple, correspondant à la lettre code et au NQA désiré, plutôt qu'un plan d'échantillonnage double ou multiple. S'il y a une flèche, la remonter jusqu'à la première inscription indiquant les critères d'acceptation et de rejet et utiliser la taille d'échantillon donnée à gauche de cette inscription, et non celle qui correspond à la lettre code originale.

Exemple 4.1 Choix d'un plan d'échantillonnage à l'aide de la norme ISO/2859-1

Prenons l'exemple d'une inspection lot par lot de poudre de lait écrémé où les lots sont composés de 250 sacs. Nous ferons une inspection normale de niveau II de la caractéristique à étudier, selon un NQA de 0.40% d'éléments défectueux.

D'après le Tableau 5A de l'Appendice, la lettre code correspondant à un lot de taille 250 est G.

Si l'on opte pour un plan d'échantillonnage simple et une inspection normale, on utilisera le Tableau 5B de l'Appendice. Dans la colonne 0.40 et la rangée G de ce tableau, on peut lire que $Ac = 0$, $Re = 1$ et la taille de l'échantillon $n = 32$. Ainsi, dans le cadre de ce plan, on doit prélever un échantillon aléatoire de 32 sacs sur les 250 du lot, inspecter chacun en prenant en considération la caractéristique à étudier, déterminer le nombre de sacs défectueux d et (i) accepter le lot si $d = 0$ et (ii) le rejeter si $d \geqslant 1$.

Par contre, si l'on opte pour un plan d'échantillonnage double dans les mêmes conditions, on utilisera le Tableau 5C de l'Appendice; on constatera que les tailles cumulatives de l'échantillon sont $n_1 = n_2 = 20$ et qu'un point (•) tient lieu d'inscription dans la colonne 0.40. Le point (•) indique deux possibilités: on peut (i) utiliser le plan d'échantillonnage simple correspondant ou (ii) avoir recours au plan d'échantillonnage double donné plus bas, le cas échéant, dans le Tableau 5C de l'Appendice. Si l'on choisit la seconde possibilité, on constate que $Ac_1 = 0$, $Re_1 = 2$, $Ac_2 = 1$ et $Re_2 = 2$. À gauche de ces inscriptions, on peut lire que $n_1 = n_2 = 80$. Il est très important de remarquer que, dans ce cas, on n'utilise pas les tailles d'échantillon $n_1 = n_2 = 20$ qui servent habituellement à la lettre G. Le plan consistera donc à prélever un échantillon aléatoire de 80 sacs dans les 250 du lot, à inspecter chacun d'eux, à déterminer le nombre de sacs défectueux d_1 et (i) à accepter le lot si $d_1 = 0$, (ii) à le rejeter si $d_1 \geqslant 2$ ou (iii) à prélever un autre échantillon du lot si $d_1 = 1$. Supposons, s'il faut avoir recours à la dernière option, que ce second échantillon de 80 sacs sur les 170 restants au sein du lot donne d_2 sacs défectueux. On acceptera le lot si $d_1 + d_2 = 1$ et on le rejettera si $d_1 + d_2 \geqslant 2$.

Remarque: Si l'on avait utilisé un plan d'échantillonnage double pour une inspection réduite plutôt que le plan ci-dessus, les critères d'acceptation et de rejet ne seraient pas consécutifs, de fait, $Ac_2 = 0$ et $Re_2 = 2$. Tout porte à croire qu'aucune décision n'aurait pu être prise si les résultats de l'inspection de deux échantillons avaient donné $d_1 + d_2 = 1$. Ce

cas est traité dans la Section 11.1.4 de la norme. On y apprend que, si cette situation se produit, le lot doit être accepté parce que le niveau de qualité précédent était excellent par rapport au NQA; il faut toutefois être conscient qu'il se peut que le niveau de qualité ait fléchi. Par conséquent, il est préférable d'abandonner l'inspection réduite et sa courbe CE plutôt "indulgente" et de revenir à une inspection normale en tenant compte des mêmes conditions.

4.7 Autres procédures d'échantillonnage de réception

Sans faire une description détaillée des autres procédures d'échantillonnage de réception, nous donnerons les grandes lignes des principes qui sous-tendent certaines des procédures les plus importantes. Nous conseillons au lecteur qui veut avoir plus d'explications concernant ces méthodes de consulter la bibliographie.

4.7.1 Plans d'échantillonnage continu

Les plans d'échantillonnage continu s'appliquent à un procédé de production qui ne comporte pas de "lots" distincts. On s'en sert généralement sur les convoyeurs, mais on peut également les utiliser pour toute opération continue où le but n'est pas de former des lots du produit en vue d'une inspection.

Afin d'avoir une vue d'ensemble de l'échantillonnage continu, prenons l'exemple d'un flot continu d'un produit comprenant 3% d'éléments défectueux. On commence par inspecter le produit en classant chaque élément de suite comme étant défectueux ou non défectueux. Si O représente un élément non défectueux et X, un élément défectueux, les résultats de l'inspection pourraient se présenter comme suit:

OOX OOOOOX OOOX OOOOOOOOOOX
OOOOOOOOX OOOOOOOOOOOOOOOX OOOO

Le nombre d'éléments non défectueux qui s'intercalent entre deux éléments défectueux consécutifs, que l'on appelle "écartement des défectueux (s)", suit généralement une séquence susceptible de faire l'objet d'une prévision probabiliste. Si un produit comprend plus de 3% d'éléments défectueux, ces éléments seront plus nombreux et l'écartement s aura tendance à se rétrécir. En revanche, si le produit comprend moins de 3% d'éléments défectueux, il y aura moins d'éléments défectueux dans la séquence et l'écartement s sera plus grand. Comme les produits de qualité différente donnent différentes séquences de s, on peut établir un "critère d'acceptation" défini sur la base de s qui rejettera plus d'éléments du produit de "mauvaise" qualité et en acceptera plus qui sont de "bonne" qualité, la définition de bon et de mauvais pouvant varier selon les applications.

On se sert du plan d'échantillonnage continu de base, appelé PEC-1, en spécifiant un "intervalle de référence (i)" qui correspond à un paramètre fixe d'un plan d'échantillonnage continu donné, indiquant le nombre d'éléments consécutifs

à inspecter, et qui ne doivent pas comporter de défauts, avant de pouvoir soumettre le processus à un échantillonnage aléatoire régulier.

Les éléments consécutifs sont d'abord inspectés au complet (100%) jusqu'à ce que l'intervalle de référence soit atteint, après quoi une fraction (f) seulement des éléments sont inspectés; on choisit alors les éléments un à un du flot de produit de manière à obtenir un échantillon sans biais de la fraction f. Un schéma opérationnel simple de PEC-1 est illustré à la Figure 4.2.

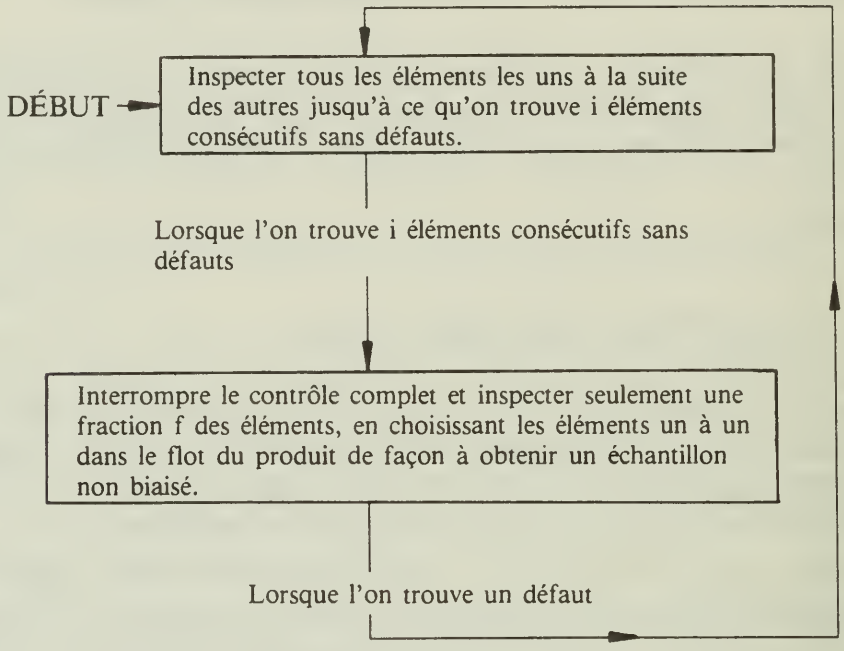


Figure 4.2 Schéma opérationnel d'un plan d'échantillonnage continu: PEC-1

Si l'on trouve un élément défectueux pendant l'échantillonnage régulier, on revient immédiatement à une inspection complète jusqu'à ce que le critère de qualification pour ce type d'échantillonnage, c'est-à-dire, l'intervalle de référence i, soit atteint à nouveau. Pour faire un échantillonnage continu, on se sert généralement d'un plan QMS, où il y a alternance des périodes d'inspection complète et des périodes d'échantillonnage aléatoire régulier. On détermine la LQMS en fonction des valeurs de i et de f spécifiées. Pour plus de renseignements concernant les procédures d'échantillonnage continu, le lecteur est invité à consulter l'ouvrage de M. Stephens (1979).

4.7.2 Plans d'échantillonnage successif partiel: ISO/2859-3

L'échantillonnage successif partiel est une procédure d'échantillonnage de réception où certains lots d'une série sont acceptés sans inspection (sauf pour certaines vérifications au hasard) quand les résultats de l'échantillonnage d'un nombre fixe de lots les précédant immédiatement répondent aux critères spécifiés. Cette pro-

cédure vise à réduire le temps consacré à l'inspection de produits soumis par des fournisseurs qui ont démontré leur aptitude à maîtriser de façon efficace toutes les facettes de la qualité des produits, et qui fabriquent immanquablement des produits de qualité supérieure.

Ces plans ne s'appliquent qu'à des séries continues de lots et ne conviennent pas à des lots isolés. Les lots à inspecter sont choisis au hasard en fonction d'une fréquence établie, appelée la "fréquence de l'échantillonnage successif partiel". Une fréquence de 1 lot dans 2, par exemple, signifie que la proportion moyenne de lots inspectés sur une longue période est de cinquante pour-cent. Le lecteur consultera la Norme internationale ISO/2859-3 pour connaître la façon de déterminer la fréquence de l'échantillonnage successif partiel ainsi que la taille de l'échantillon.

4.7.3 Plans d'échantillonnage progressif par attributs: ISO/8422

Dans un plan d'échantillonnage progressif par attributs, on choisit des éléments au hasard qu'on inspecte un à un, et on retient le compte cumulatif des éléments non-conformes (ou des non-conformités). Après avoir inspecté chaque élément, on se sert de ce compte cumulatif pour déterminer si l'on a suffisamment d'information pour décider d'accepter ou de rejeter le lot.

Si, d'après le compte cumulatif obtenu à un certain moment de l'inspection, le risque d'accepter un lot de qualité peu satisfaisante (le risque à la consommation) est suffisamment faible, le lot sera jugé acceptable et l'échantillonnage de ce lot prendra fin.

Par contre, si d'après le compte cumulatif le risque de non-acceptation d'un lot de qualité satisfaisante (le risque à la production) est suffisamment faible, le lot sera jugé non acceptable et l'échantillonnage de ce lot prendra fin.

Si le compte cumulatif ne permet de prendre aucune décision, on doit alors inspecter un autre élément du produit. L'inspection se poursuivra jusqu'à ce qu'on ait accumulé suffisamment d'information pour prendre une décision finale envers le lot. Pour plus de renseignements, consultez la Norme ISO/8422: Plans d'échantillonnage progressif pour le contrôle par attributs.

CHAPITRE 5

CONTRÔLE STATISTIQUE DU PROCESSUS

5.1 Introduction

Dans le domaine de la fabrication, on compte habituellement sur le service de production pour fabriquer le produit et sur le personnel chargé de la maîtrise de la qualité pour inspecter le produit final et éliminer les éléments qui ne répondent pas aux spécifications. Cette stratégie de détection est peu économique parce qu'elle suppose un contrôle après coup, lorsque le produit inacceptable est déjà fait. Au lieu de cela, il est beaucoup plus avantageux d'adopter une stratégie préventive pour éviter les pertes en ne fabriquant pas au départ un produit inutilisable. Pour ce faire, on recueille des données sur le processus et on les analyse de façon à pouvoir prendre des mesures qui s'appliquent au processus lui-même. Cette fonction est celle des méthodes de **contrôle statistique du processus (CSP).**

Le contrôle statistique du processus a pour but de permettre d'établir un processus et de le maintenir à un niveau acceptable et stable afin que les produits et les services soient continuellement conformes aux spécifications. L'outil statistique qu'on utilise le plus souvent à cette fin est le **graphique de contrôle** qui sert à présenter et à comparer graphiquement des informations. Le graphique s'appuie sur une séquence d'échantillons représentant l'état actuel d'un processus par rapport à des limites établies une fois qu'on a pris en considération la variabilité inhérente au processus. Le graphique de contrôle permet tout d'abord de déterminer si un processus a atteint un état de contrôle statistique, ou est toujours dans un état de contrôle statistique, selon le bon niveau spécifié, puis d'obtenir et de maintenir l'état de contrôle et l'uniformité idéale des caractéristiques importantes des produits et des services en consignant continuellement la qualité du produit en cours de fabrication.

C'est D[r] Walter Shewhart qui, en 1924, a été le premier à suggérer d'utiliser le graphique de contrôle comme moyen d'appliquer graphiquement les principes de la signification statistique au contrôle de la production. La théorie des graphiques de contrôle repose sur la reconnaissance de deux types de variabilité. Le premier dépend de "**causes dues au hasard**" ou de "**causes communes**"; il s'agit de la variabilité aléatoire. L'élimination ou la correction des causes communes fait appel à une répartition des ressources par la direction dans le but d'améliorer le processus et le système. Le second type de variabilité représente une transformation concrète du processus qui peut être imputable à des causes identifiables ne faisant pas partie du processus et pouvant, du moins en théorie, être éliminées. Ces causes identifiables s'appellent aussi "**causes assignables**" ou "**causes spéciales**" de la variation. Elles peuvent être attribuables à un manque d'uniformité des matériaux, à un manque d'uniformité à l'exécution ou aux procédures, à un outil brisé, ou au fonctionnement irrégulier du matériel de fabrication ou d'essai.

Les graphiques de contrôle permettent de détecter dans les données des régimes de variation anormaux qui découlent de la répétition des processus, et fournissent des critères pour déceler un manque de contrôle statistique. Un processus est dans un état de **contrôle statistique** lorsque la variabilité est uniquement attribuable à des causes aléatoires ou communes. Une fois que ce niveau acceptable de variation est déterminé, on suppose que tout écart par rapport à ce niveau est dû à des causes assignables qu'il faudrait identifier pour les éliminer ou en réduire le nombre.

5.2 Types de graphiques de contrôle

Voici les types de graphiques de contrôle qu'on utilise le plus souvent pour étudier le contrôle des processus:

- **Système de graphiques de Shewhart**
  - — Variables: Graphique de la moyenne ($\overline{X}$), graphique de l'étendue (R) ou de l'écart-type (s)
  - — Attributs: Graphique du pourcentage d'éléments défectueux (p), graphique du nombre d'éléments défectueux (np), graphique du nombre de défauts (c), graphique du nombre de défauts par unité (u)
  - — Graphiques des valeurs individuelles (X) et des étendues mobiles (R)
  - — Graphique de la médiane (Me) et graphique de l'étendue (R)
- **Graphiques à somme cumulée**
- **Graphiques de contrôle de réception**

Notre analyse portera seulement sur les graphiques $\overline{X}$ et R des variables de Shewhart, le graphique p des attributs et les graphiques à somme cumulée.

5.3 Système de graphiques de Shewhart

Pour construire un graphique de contrôle de Shewhart, on doit se servir des données obtenues des échantillons prélevés à intervalles réguliers dans le processus. Les intervalles peuvent se définir en termes de temps (toutes les heures, par exemple) ou de quantité (chaque lot, par exemple) et déterminent les **sous-groupes** ou échantillons choisis dans le processus. On établit, pour chaque sous-groupe, une caractéristique ou plus, comme la moyenne du sous-groupe, $\overline{X}$, et l'étendue du sous-groupe, R, ou son écart-type, s. Le graphique est constitué des valeurs d'une caractéristique de sous-groupe donnée indiquées vis-à-vis les numéros des sous-groupes, d'une **ligne centrale (LC)** qui est tracée en fonction d'une valeur de référence, et de deux limites de contrôle déterminées statistiquement, une de chaque côté de la ligne centrale, qu'on appelle la **limite supérieure de contrôle (LSC)** et la **limite inférieure de contrôle (LIC)** (se reporter à la Figure 5.1).

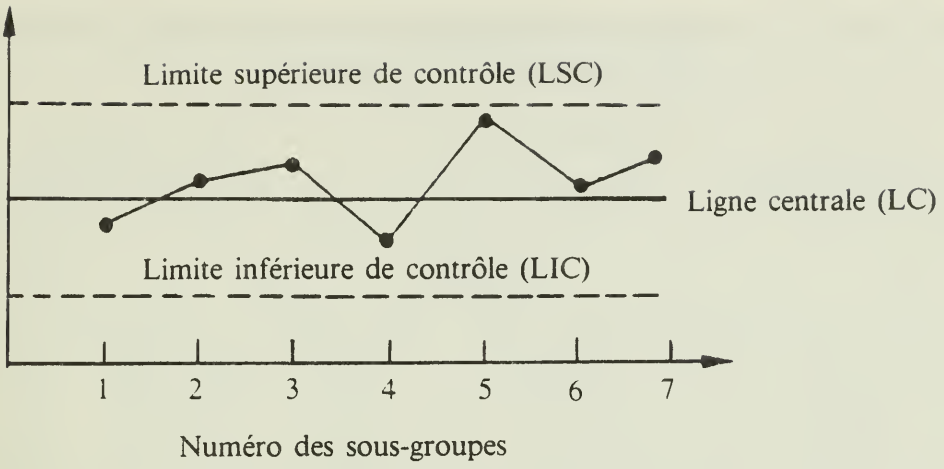


Figure 5.1 Profil d'un graphique de contrôle

Shewhart a proposé de fixer les limites des graphiques de contrôle à une distance de 3σ de chaque côté de la ligne centrale, où σ représente l'écart-type intérieur de la population des sous-groupes estimé à partir des étendues ou des écarts-types des échantillons. Cela signifie qu'environ 99.7% des valeurs des sous-groupes tomberont à l'intérieur des limites de contrôle, dans la mesure où le processus est dans un état de contrôle statistique. Cela signifie également qu'il y a environ 0.3% de chance (ou trois chances sur mille, en moyenne) qu'une valeur tombe par erreur en dehors des limites lorsque le processus est bien centré et dans un état de contrôle. Il est tellement peu probable qu'une violation de ces limites soit imputable au hasard plutôt que révélatrice d'une transformation concrète du processus que, si un point se situe en dehors des limites, des mesures devraient être prises. C'est d'ailleurs pour cette raison que ces limites de contrôle de 3σ sont parfois appelées **limites de tolérance.** On trace souvent aussi des limites de 2σ (appelées **limites d'alerte)** pour mettre en garde contre une situation de perte de contrôle imminente lorsqu'une valeur ne se situe pas dans ces limites.

Lorsqu'une valeur représentée graphiquement tombe en dehors de l'une ou l'autre des limites de contrôle ou qu'une série de valeurs présente une configuration inhabituelle, l'état de contrôle statistique n'est plus acceptable. Il faut alors chercher la(les) cause(s) assignable(s) et arrêter ou corriger le processus. Une fois qu'on a déterminé et éliminé la cause, le processus peut se poursuivre. On pourra établir les paramètres permanents d'un graphique de contrôle lorsqu'on aura réussi à obtenir un graphique de contrôle qui présente un état de contrôle pour un nombre raisonnable de sous-groupes.

5.4 Graphiques de contrôle de Shewhart: Formules et facteurs

Le Tableau 5.1 indique les formules requises pour tracer les lignes d'un graphique de contrôle des variables, des attributs et des valeurs individuelles.

TABLEAU 5.1: Formules des limites de contrôle des graphiques de Shewhart

Type	Ligne centrale (LC)	Limite supérieure de contrôle (LSC)	Limite inférieure de contrôle (LIC)
		Variables	
$\bar{X}$	$\bar{\bar{X}}$	$\bar{\bar{X}} + A_2\bar{R}$ ou $\bar{\bar{X}} + A_3\bar{s}$	$\bar{\bar{X}} - A_2\bar{R}$ ou $\bar{\bar{X}} - A_3\bar{s}$
R	$\bar{R}$	$D_4\bar{R}$	$D_3\bar{R}$
s	$\bar{s}$	$B_4\bar{s}$	$B_3\bar{s}$
		Attributs	
p	$\bar{p}$	$\bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$	$\bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$
np	$n\bar{p}$	$n\bar{p} + 3\sqrt{n\bar{p}(1-\bar{p})}$	$n\bar{p} - 3\sqrt{n\bar{p}(1-\bar{p})}$
c	$\bar{c}$	$\bar{c} + 3\sqrt{\bar{c}}$	$\bar{c} - 3\sqrt{\bar{c}}$
u	$\bar{u}$	$\bar{u} + 3\sqrt{\frac{\bar{u}}{n}}$	$\bar{u} - 3\sqrt{\frac{\bar{u}}{n}}$
		Valeurs individuelles	
X	$\bar{X}$	$\bar{X} + E_2\bar{R}$	$\bar{X} - E_2\bar{R}$

Remarque: 1. Les valeurs des facteurs A_2, A_3, D_3, D_4, B_3 et B_4 sont données dans le Tableau 5.2 ci-dessous pour différentes valeurs de la taille n des sous-groupes.

2. Les symboles $\bar{\bar{X}}$, $\bar{R}$ et $\bar{s}$ représentent la moyenne des moyennes, des étendues et des écarts-types des sous-groupes respectivement.

3. Dans le cas des graphiques des valeurs individuelles où une seule observation par sous-groupe est présentée, on obtient une mesure de variabilité en calculant l'étendue mobile de deux observations. Une étendue mobile correspond à la différence absolue entre des paires de mesures successives dans une série, c.-à-d., la différence absolue entre la première et la deuxième mesure, puis entre la deuxième et la troisième mesure, et ainsi de suite. On peut alors calculer l'étendue mobile moyenne, $\bar{R}$, à partir de ces données et s'en servir pour construire des graphiques de contrôle des valeurs individuelles. La valeur du facteur E_2 est obtenue par $3/d_2$, où les valeurs de d_2 sont indiquées dans le Tableau 5.2 pour la taille n des sous-groupes. Prenons l'exemple d'une étendue mobile de deux observations; n est donc égal à 2 et, par conséquent, $E_2 = 3/1.128 = 2.66$.

TABLEAU 5.2: Facteurs des graphiques de contrôle des variables de Shewhart

Nombre d'observations dans un sous-groupe (n)	A_2	A_3	D_3	D_4	B_3	B_4	d_2
2	1.880	2.659	—	3.267	—	3.267	1.128
3	1.023	1.954	—	2.574	—	2.568	1.693
4	0.729	1.628	—	2.282	—	2.266	2.059
5	0.577	1.427	—	2.114	—	2.089	2.326
6	0.483	1.287	—	2.004	0.030	1.970	2.534
7	0.419	1.182	0.076	1.924	0.118	1.882	2.704
8	0.373	1.099	0.136	1.864	0.185	1.815	2.847
9	0.337	1.032	0.184	1.816	0.239	1.761	2.970
10	0.308	0.975	0.223	1.777	0.284	1.716	3.078
11	0.285	0.927	0.256	1.744	0.321	1.679	3.173
12	0.266	0.886	0.283	1.717	0.354	1.646	3.258
13	0.249	0.850	0.307	1.693	0.382	1.618	3.336
14	0.235	0.817	0.328	1.672	0.406	1.594	3.407
15	0.223	0.789	0.347	1.653	0.428	1.572	3.472

5.5 Construction des graphiques de contrôle de Shewhart

Nous donnons ci-après, en tant qu'exemple, la marche à suivre typique pour construire un graphique $\overline{X}$ et un graphique R. La même marche à suivre peut s'appliquer à tous les autres types de graphiques de contrôle.

1. Recueillir des données, sous-groupe par sous-groupe, en prélevant 20 à 25 sous-groupes (k), ayant chacun une taille de 4 ou 5 (n), et en mesurant la caractéristique désirée. On veillera dans la classification des observations en sous-groupes à ce que la variation au sein d'un sous-groupe semble uniquement attribuable à des causes aléatoires et la variation entre les sous-groupes semble imputable à des causes assignables que le graphique de contrôle a pour objet de détecter.

2. Calculer, pour chaque sous-groupe, la moyenne $\overline{X}$ et l'étendue R.

3. Calculer la moyenne globale de toutes les observations, $\overline{\overline{X}}$, c.-à-d., la moyenne de toutes les moyennes des sous-groupes, et la moyenne de toutes les étendues des sous-groupes, $\overline{R}$.

4. Tracer, sur un papier quadrillé ou sur une feuille appropriée, un graphique $\overline{X}$ ainsi qu'un graphique R. L'échelle verticale à gauche sert à représenter la variable $\overline{X}$ ou R, selon le cas, et l'échelle horizontale, le numéro des sous-groupes. Vis-à-vis les numéros des sous-groupes, indiquer à l'aide de points, sur le graphique des moyennes, les valeurs de $\overline{X}$ calculées et, sur le graphique

des étendues, les valeurs de R. Tracer sur le graphique approprié des lignes horizontales ininterrompues pour représenter $\overline{\overline{X}}$ et $\overline{R}$, déterminant ainsi la position des lignes centrales.

5. Indiquer les limites de contrôle. Tracer sur le graphique $\overline{X}$ deux lignes pointillées horizontales aux niveaux $\overline{\overline{X}} \pm A_2\overline{R}$ et, sur le graphique R, une ligne pointillée horizontale au niveau $D_3\overline{R}$ et une autre au niveau $D_4\overline{R}$, où la valeur de A_2, D_3 et D_4 dépend de la valeur de n, le nombre d'observations dans un sous-groupe, et se retrouvent au Tableau 5.2. Chaque fois que la valeur de n est inférieure à 7, il n'est pas nécessaire d'indiquer la LIC sur le graphique R puisqu'on considère alors D_3 comme étant égal à zéro.

6. Tracer d'abord le graphique R en reliant les points consécutifs par un trait. Vérifier s'il y a des points qui tombent en dehors des limites de contrôle, signalant ainsi une perte de contrôle, et si le graphique présente une configuration ou une tendance inhabituelle. Faire, pour chaque point indiquant une cause assignable dans l'étendue des données, une analyse du processus afin de déterminer la cause; corriger la situation et voir à ce qu'elle ne se reproduise pas.

7. Exclure tous les sous-groupes rattachés à une cause assignable identifiée; puis, recalculer et tracer la nouvelle étendue moyenne, $\overline{R}$, avec ses limites de contrôle révisées. Vérifier si tous les points pour l'étendue rendent compte d'un état de contrôle statistique lorsqu'ils sont comparés aux nouvelles limites, et répéter, au besoin, les étapes de l'identification, de la correction et du calcul.

8. Tous les sous-groupes retirés du graphique R parce qu'on pouvait les rattacher à des causes assignables devraient également être exclus du graphique $\overline{X}$. On pourra se servir des valeurs révisées de $\overline{R}$ et de $\overline{\overline{X}}$ pour recalculer les limites de contrôle d'essai pour les moyennes, $\overline{\overline{X}} \pm A_2\overline{R}$.

9. On considère que la dispersion du processus, c.-à-d., la variation à l'intérieur des sous-groupes, est stable si les étendues sont dans un état de contrôle statistique. On peut alors analyser les moyennes pour voir si la valeur centrale du processus fluctue parfois.

10. Tracer maintenant le graphique $\overline{X}$ et vérifier s'il y a des points qui tombent en dehors des limites de contrôle, signalant ainsi une perte de contrôle, et si le graphique présente une configuration ou une tendance inhabituelle. Analyser, comme dans le cas du graphique R, toute situation de perte de contrôle et prendre des mesures correctives et préventives. Exclure tous les sous-groupes correspondant à une perte de contrôle et auxquels on peut rattacher des causes assignables; recalculer et tracer la nouvelle moyenne de production, $\overline{\overline{X}}$, avec ses limites de contrôle révisées et s'assurer que le processus est dans un état de contrôle statistique en répétant, au besoin, les étapes de l'identification, de la correction et du calcul.

11. Lorsque les valeurs de l'étendue et de la moyenne tombent constamment à l'intérieur des limites de contrôle d'essai, reporter ces limites à des périodes futures. Ces limites serviraient alors à contrôler continuellement le processus et permettraient aux personnes responsables de réagir promptement aux indications de perte de contrôle révélées par le graphique $\overline{X}$ ou R.

5.6 Maîtrise et aptitude du processus

Un processus est considéré comme stable si les points correspondant aux données des sous-groupes sur le graphique tombent à l'intérieur des limites de contrôle. On peut identifier une situation qui n'est pas dans un état de contrôle en se fondant sur les critères suivants:

- un point qui tombe à l'extérieur des limites de contrôle
- sept (7) points consécutifs qui se trouvent du même côté de la ligne centrale
- sept (7) points consécutifs qui se déplacent progressivement vers le haut ou vers le bas
- toute autre configuration évidemment non aléatoire comme des cycles, un changement graduel de niveau, un groupement, une interaction, un arrangement systématique, des tendances, etc.

Après avoir établi l'état de contrôle statistique d'un processus, on étudie l'aptitude de celui-ci à répondre aux spécifications. **L'aptitude du processus** représente la performance du processus même que l'on peut évaluer une fois qu'on a identifié, analysé et corrigé toutes les causes spéciales des situations de perte de contrôle mises en évidence par les graphiques $\bar{X}$ et R, et qu'on a fait en sorte que la situation ne se reproduise plus.

D'une façon générale, l'aptitude du processus s'exprime sous forme d'indice, appelé **indice de l'aptitude du processus, IAP (ou C_p)**, comme suit:

$$\text{IAP} = \frac{\text{Tolérance spécifiée}}{\text{Aptitude du processus}} = \frac{\text{LST} - \text{LIT}}{6\sigma}$$

où

LST représente la limite supérieure de tolérance,

LIT représente la limite inférieure de tolérance, et

σ a sa valeur estimée à partir de la variabilité au sein des sous-groupes, mesurée comme $\frac{\bar{R}}{d_2}$.

On remarquera que dans le cas des graphiques de contrôle p et np, l'aptitude du processus correspond à la proportion moyenne d'éléments qui sont conformes aux spécifications. Ainsi, l'aptitude est égale à $1 - \bar{p}$. Les graphiques de contrôle c et u ne permettent pas d'exprimer l'aptitude du processus de la même façon. Les attributs $\bar{c}$ et $\bar{u}$ sont alors les mesures respectives de la performance du processus.

Exemple 5.1 Graphique $\bar{X}$ et graphique R

Le Tableau 5.3 nous donne les mesures de la teneur en humidité d'échantillons de poudre de lait écrémé. On prend chaque demi-heure, pendant 10 heures, quatre

mesures indépendantes pour obtenir k = 20 sous-groupes (échantillons) de taille n = 4. Le Tableau 5.3 indique également les moyennes et les étendues des sous-groupes. Les limites supérieure et inférieure de tolérance sont respectivement de 0.219% et de 0.125%. Il s'agit d'évaluer la performance du processus et de contrôler statistiquement la valeur centrale et la dispersion du processus pour que celui-ci réponde aux spécifications.

TABLEAU 5.3: Données de production sur la teneur en humidité (%) de la poudre de lait écrémé

Numéro des sous-groupes	Teneur en humidité (%)				Moyenne ($\bar{X}$)	Étendue (R)
	X_1	X_2	X_3	X_4		
1	0.1898	0.1729	0.2067	0.1898	0.1898	0.0338
2	0.2012	0.1913	0.1878	0.1921	0.1931	0.0134
3	0.2217	0.2192	0.2078	0.1980	0.2117	0.0237
4	0.1832	0.1812	0.1963	0.1800	0.1852	0.0163
5	0.1692	0.2263	0.2086	0.2091	0.2033	0.0571
6	0.1621	0.1832	0.1914	0.1783	0.1788	0.0293
7	0.2001	0.1927	0.2169	0.2082	0.2045	0.0242
8	0.2401	0.1825	0.1910	0.2264	0.2100	0.0576
9	0.1996	0.1980	0.2076	0.2023	0.2019	0.0096
10	0.1793	0.1715	0.1829	0.1961	0.1822	0.0246
11	0.2166	0.1748	0.1960	0.1923	0.1949	0.0418
12	0.1924	0.1984	0.2377	0.2003	0.2072	0.0453
13	0.1768	0.1986	0.2241	0.2022	0.2004	0.0473
14	0.1923	0.1876	0.1903	0.1986	0.1922	0.0110
15	0.1924	0.1996	0.2120	0.2160	0.2050	0.0236
16	0.1720	0.1940	0.2116	0.2320	0.2024	0.0600
17	0.1824	0.1790	0.1876	0.1821	0.1828	0.0086
18	0.1812	0.1585	0.1699	0.1680	0.1694	0.0227
19	0.1700	0.1567	0.1694	0.1702	0.1666	0.0135
20	0.1698	0.1664	0.1700	0.1600	0.1666	0.0100

Dans cet exemple,

$$\bar{\bar{X}} = \frac{\Sigma \bar{X}}{k} = \frac{3.8480}{20} = 0.1924 \ .$$

$$\bar{R} = \frac{\Sigma R}{k} = \frac{0.5734}{20} = 0.0287 \ .$$

Il faut tout d'abord tracer un graphique R et évaluer l'état de contrôle.

Graphique R:

$$\text{Ligne centrale} = \overline{R} = 0.0287$$

$$\text{LSC} = D_4\overline{R} = 2.282 \times 0.0287 = 0.0655$$

$$\text{LIC} = D_3\overline{R} = 0 \times 0.0287 = -$$

Remarque: Si la valeur de n est inférieure à 7, la LIC ne figure pas sur le graphique.

Les valeurs des facteurs de multiplication D_3 et D_4 sont tirées du Tableau 5.2 pour $n = 4$. Étant donné que les valeurs de R du Tableau 5.3 se situent toutes dans les limites de contrôle du graphique R, celui-ci indique que la dispersion est dans un état de contrôle statistique. On peut donc se servir de la valeur de $\overline{R}$ pour calculer les limites de contrôle du graphique $\overline{X}$.

Graphique $\overline{X}$:

$$\text{Ligne centrale} = \overline{\overline{X}} = 0.1924$$

$$\text{LSC} = \overline{\overline{X}} + A_2\overline{R} = 0.1924 + (0.729 \times 0.0287) = 0.2133$$

$$\text{LIC} = \overline{\overline{X}} - A_2\overline{R} = 0.1924 - (0.729 \times 0.0287) = 0.1715$$

La valeur du facteur A_2 est tirée du Tableau 5.2 pour $n = 4$. Les graphiques de contrôle $\overline{X}$ et R sont représentés à la Figure 5.2. Un examen du graphique $\overline{X}$ nous indique que les trois derniers points dénotent une perte de contrôle du processus qui est due à certaines causes de variation assignables. Si les limites de contrôle avaient été calculées à partir de données précédentes, il y aurait eu lieu de prendre des mesures pour le sous-groupe 18.

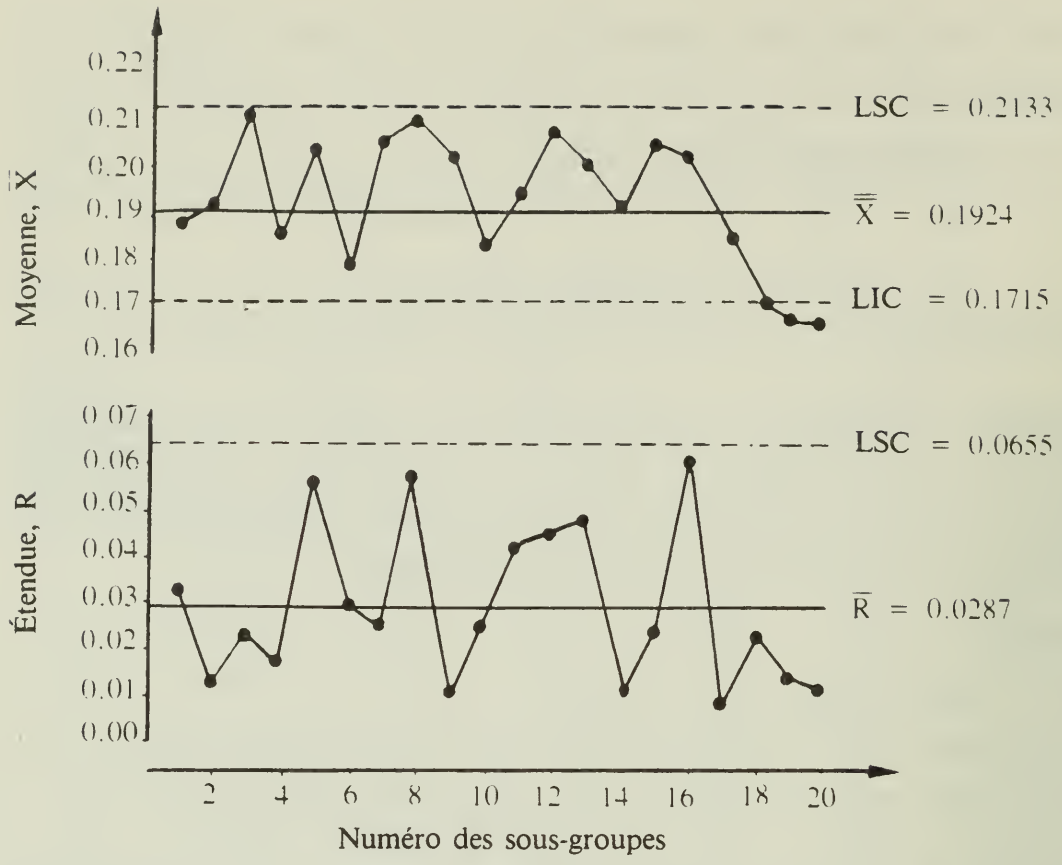


Figure 5.2 Graphiques de la moyenne et de l'étendue pour les données du Tableau 5.3

Des mesures correctives appropriées sont prises à ce moment-ci pour éliminer les causes assignables et empêcher leur récurrence. On reconstruit les graphiques à l'aide des limites de contrôle révisées en rejetant les points révélateurs d'une perte de contrôle, c.-à-d., les valeurs des sous-groupes 18, 19 et 20. On procède comme suit pour recalculer les valeurs de $\bar{\bar{X}}$ et de $\bar{R}$ et tracer les nouvelles lignes des graphiques de contrôle:

$$\text{Valeur révisée de } \bar{\bar{X}} = \frac{\Sigma \bar{X}}{k} = \frac{3.3454}{17} = 0.1968$$

$$\text{Valeur révisée de } \bar{R} = \frac{\Sigma R}{k} = \frac{0.5272}{17} = 0.0310$$

Graphique $\bar{X}$ révisé:

$$\text{Ligne centrale} = \bar{\bar{X}} = 0.1968$$

$$\text{LSC} = \bar{\bar{X}} + A_2\bar{R} = 0.1968 + (0.729 \times 0.031) = 0.2194$$

$$\text{LIC} = \bar{\bar{X}} - A_2\bar{R} = 0.1968 - (0.729 \times 0.031) = 0.1742$$

Graphique R révisé:

Ligne centrale = $\overline{R}$ = 0.0310

LSC = $D_4\overline{R}$ = 2.282 × 0.0310 = 0.0707

LIC = $D_3\overline{R}$ = 0 × 0.0310 = –

Remarque: Si la valeur de n est inférieure à 7, la LIC ne figure pas sur le graphique.

Les graphiques de contrôle révisés sont illustrés à la Figure 5.3 et indiquent que le processus est dans un état de contrôle statistique pour ce qui est de la valeur centrale et de la dispersion.

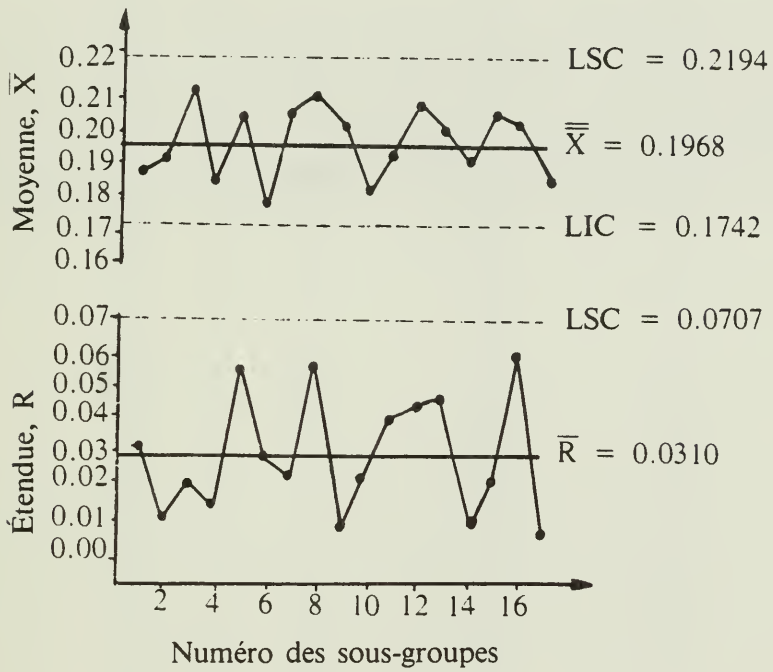


Figure 5.3 Graphiques $\overline{X}$ et R révisés des données du Tableau 5.3

Comme le processus présente un état de contrôle statistique depuis la révision des limites de contrôle, on peut donc évaluer l'indice d'aptitude du processus de la façon suivante:

$$IAP = \frac{\text{Tolérance spécifiée}}{\text{Aptitude du processus}} = \frac{LST - LIT}{6\sigma}$$

où l'estimation de $\sigma = \overline{R}/d_2 = \frac{0.031}{2.059} = 0.0151$.

Par conséquent:

$$IAP = \frac{0.219 - 0.125}{6 \times 0.0151} = \frac{0.094}{0.0906} = 1.04 .$$

Le Tableau 5.2 donne la valeur de la quantité d_2 pour $n = 4$. Puisque l'IAP est supérieur à 1, l'aptitude du processus semble acquise. Toutefois, en y regardant de plus près, on peut constater que le processus n'est pas bien centré par rapport aux spécifications et, pour cette raison, environ 11.8% des mesures excèdent la limite supérieure des spécifications. Il importe donc, avant d'établir les paramètres permanents des graphiques de contrôle, de bien centrer le processus tout en maintenant celui-ci dans un état de contrôle statistique.

Exemple 5.2 Graphique de la fraction d'éléments non-conformes (p)

On a décidé d'installer dans une entreprise de conditionnement des aliments un graphique de la fraction d'éléments non-conformes p dans le but de contrôler la performance de l'étiqueteuse de conserves de produits alimentaires. Des données ont été recueillies et analysées pendant 26 jours ouvrables. On a prélevé à la fin de chaque jour un échantillon aléatoire de boîtes et on a examiné chacune d'elles pour voir si elle était mal étiquetée. Le Tableau 5.4 indique les résultats.

TABLEAU 5.4: Contrôle de la performance de l'étiqueteuse

Numéro des sous-groupes	Nombre de boîtes inspectées	Nombre d'éléments non-conformes	Fraction d'éléments non-conformes	LSC	LIC
1	158	11	0.070	0.117	0.003
2	140	11	0.079	0.120	
3	140	8	0.057	0.120	
4	155	6	0.039	0.117	0.003
5	160	4	0.025	0.116	0.004
6	144	7	0.049	0.119	0.001
7	139	10	0.072	0.120	
8	151	11	0.073	0.118	0.002
9	163	9	0.055	0.116	0.004
10	148	5	0.034	0.119	0.001
11	150	2	0.013	0.118	0.002
12	153	7	0.046	0.118	0.002
13	149	7	0.047	0.118	0.002
14	145	8	0.055	0.119	0.001
15	160	6	0.038	0.116	0.004
16	165	15	0.091	0.115	0.005
17	136	18	0.132	0.121	
18	153	10	0.065	0.118	0.002
19	150	9	0.060	0.118	0.002
20	148	5	0.034	0.119	0.001
21	135	0	0.000	0.121	
22	165	12	0.073	0.115	0.005
23	143	10	0.070	0.120	0.000
24	138	8	0.058	0.121	
25	144	14	0.097	0.119	0.001
26	161	20	0.124	0.116	0.004
Total	3893	233			

On peut également lire dans le Tableau 5.4 les valeurs de la fraction d'éléments non-conformes, p, calculées pour chaque sous-groupe. Le calcul de la fraction moyenne mensuelle d'éléments non-conformes se fait comme suit:

$$\bar{p} = \frac{\text{Nombre total d'éléments non-conformes}}{\text{Nombre total de boîtes inspectées}} = \frac{233}{3893} = 0.060 .$$

Étant donné que la taille des sous-groupes varie, on calcule séparément les valeurs de la LSC et de la LIC de chaque sous-groupe comme suit:

$$\bar{p} \pm 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

où n correspond à la taille du sous-groupe. Ces valeurs figurent également sur le Tableau 5.4.

Représenter graphiquement les valeurs de la LSC et de la LIC de chaque sous-groupe exige, comme on peut le constater, beaucoup de temps. On peut toutefois remarquer à la lecture du Tableau 5.4 que la fraction d'éléments non-conformes des sous-groupes 17 et 26 dépassent leurs limites supérieures de contrôle. Ces deux sous-groupes sont rejetés parce qu'ils sont sujets à des variations différentes de celles des autres sous-groupes. Leur inclusion dans les calculs aurait pour effet de surestimer la moyenne du processus et ferait en sorte que les limites de contrôle ne rendraient pas compte des variations aléatoires réelles de celui-ci. Il y aurait lieu de chercher les causes de ces valeurs élevées afin de pouvoir les corriger et empêcher qu'elles ne se reproduisent. On calcule ensuite une fraction moyenne d'éléments non-conformes révisée en se servant des valeurs des 24 sous-groupes restants comme suit:

$$\bar{p} = \frac{195}{3596} = 0.054 \ .$$

Un calcul des valeurs révisées de la LSC et de la LIC pour chaque sous-groupe à l'aide de la valeur révisée de $\bar{p}$ permettrait de constater que toutes les fractions d'éléments non-conformes sont inférieures à leurs valeurs correspondantes de la LSC. Cette valeur révisée de $\bar{p}$ devient donc la fraction d'éléments non-conformes type qui servira à établir les graphiques de contrôle où la ligne centrale se retrouvera alors au niveau $\bar{p} = 0.054$.

Comme nous l'avons déjà mentionné, il faut beaucoup de temps pour représenter graphiquement les limites supérieures de contrôle de chaque sous-groupe de taille différente, et cette tâche est fastidieuse. Toutefois, comme les tailles des sous-groupes ne varient pas beaucoup par rapport à leur taille moyenne, qu'on évalue à environ 150, on peut établir le graphique p révisé en se fondant sur cette taille moyenne ($n = 150$) pour définir la valeur de la limite supérieure de contrôle. On calculera donc ainsi les lignes du graphique p révisé:

Graphique p révisé:

$$\text{Ligne centrale} = \bar{p} = 0.054$$

$$\text{LSC} = \bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 0.054 + 3 \times \sqrt{\frac{0.054 \times 0.946}{150}}$$

$$= 0.109$$

$$\text{LIC} = \bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 0.054 - 3 \times \sqrt{\frac{0.054 \times 0.946}{150}}$$

$$= -$$

Remarque: Puisque les valeurs négatives pour p sont inadmissibles, la limite inférieure de contrôle ne figure pas sur le graphique.

Le graphique p révisé est représenté à la Figure 5.4 et démontre que l'étiqueteuse est dans un état de contrôle statistique.

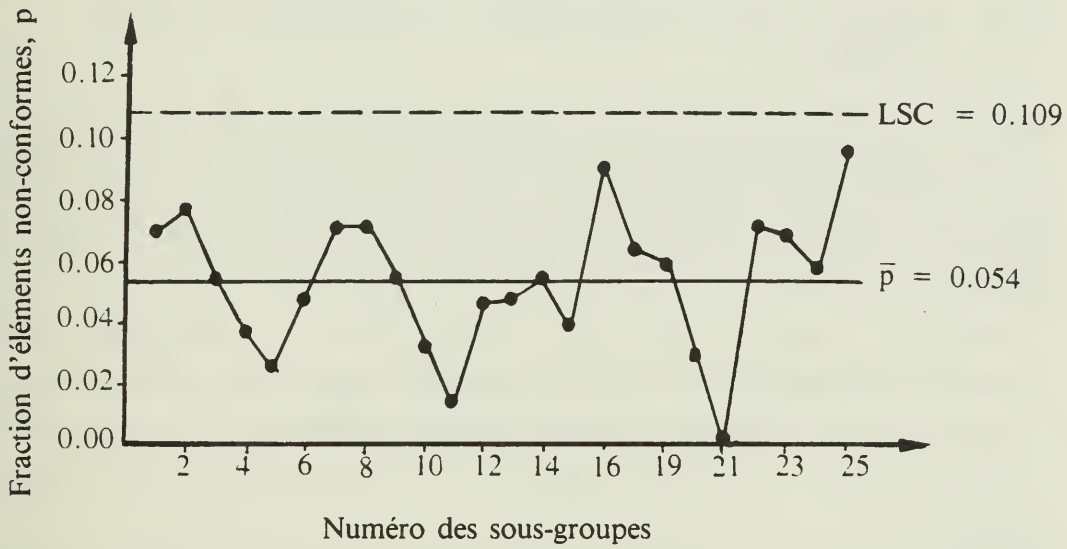


Figure 5.4 Graphique p révisé des données du Tableau 5.4

5.7 Graphiques de contrôle à somme cumulée

M. Page proposa en 1961 une alternative aux graphiques de contrôle de Shewhart, les **graphiques à somme cumulée,** dans le but d'exercer un contrôle plus étroit sur le processus grâce à la détection et à la correction plus rapides de petits écarts du processus par rapport à la valeur centrale. Le graphique de contrôle de Shewhart ne se sert que des informations du processus contenues dans le dernier point du graphique et ne tient pas compte de l'information fournie par la séquence complète de points, si ce n'est par le biais d'essais portant sur une suite de points ou en faisant appel à des limites d'alerte. Le graphique à somme cumulée incor-

pore directement l'information relative à la séquence complète de valeurs des échantillons en permettant de porter sur le graphique la somme cumulative des écarts de ces valeurs par rapport à une valeur cible préétablie.

Le graphique de Shewhart comme le graphique à somme cumulée présentent des avantages et des inconvénients. Alors que le graphique de Shewhart permet de détecter plus facilement des transformations plus importantes du processus sur une courte période, le graphique à somme cumulée offre l'avantage de détecter des changements constants à l'intérieur de la région 0.5σ à 2σ. Tout en permettant de visualiser plus rapidement les petites variations du processus et la période où ces variations se sont produites, le graphique à somme cumulée est lent à détecter des variations importantes et rend difficile l'interprétation des configurations parce que les points ne sont pas indépendants et sans corrélation. Un analyste avisé se servirait des graphiques à somme cumulée pour détecter plus rapidement de petites variations dans le processus et des graphiques de Shewhart pour analyser des données antérieures dans le but de déceler un manque de contrôle et de rétablir par la suite l'état de contrôle statistique au processus.

5.7.1 Marche à suivre pour établir un graphique à somme cumulée

Le graphique à somme cumulée est une représentation de la somme cumulative des écarts d'un processus par rapport à une valeur de contrôle ou à une valeur cible. Malgré la grande complexité des concepts mathématiques et des règles de décision qui sous-tendent ce type de graphique, la construction de celui-ci est simple.

1. Identifier la valeur de contrôle ou la valeur cible, T. La valeur cible correspond à la valeur acceptée ou prévue de la variable qui fait l'objet de l'examen.
2. Calculer les écarts de T par rapport à chaque moyenne de sous-groupe observée, c.-à-d., $\overline{X}_i - T$, où $\overline{X}_i$ correspond à la moyenne des mesures du i^e sous-groupe ($i = 1, 2, \ldots, k$).
3. Calculer et indiquer sur un papier quadrillé la somme cumulative de ces écarts,

$$S_k = \sum_{i=1}^{k} (\overline{X}_i - T) ,$$

par rapport au nombre total de sous-groupes enregistrés, k. Une convention pratique pour l'établissement de l'échelle consiste à considérer la distance entre les valeurs consécutives de somme cumulée (S_k) sur l'axe horizontal X comme une unité et la même distance sur l'axe vertical Y comme étant égale à environ 2σ unités, où la valeur estimée de σ correspond, comme dans le cas des graphiques de Shewhart, à la moyenne des écarts-types au sein des sous-groupes.

5.7.2 Règles de décision relatives aux graphiques à somme cumulée

Pour qu'un outil statistique soit utilisé par beaucoup de non-statisticiens, il est extrêmement important que cet outil se caractérise par un certain degré de simplicité et de normalisation. Les graphiques de contrôle de Shewhart qui présentent des limites de 2σ et de 3σ possèdent ces qualités, et s'ils reçoivent un si bon accueil, c'est en grande partie grâce à celles-ci. En revanche, les règles de décision exactes des graphiques à somme cumulée sont très complexes. Ces règles nécessitent la construction d'un **masque en V** qui se fonde sur les solutions approximatives d'un trajet aléatoire, les côtés du V comportant des barrières absorbantes. Le masque en V est superposé au graphique à somme cumulée et définit un intervalle de décision. Le processus sera considéré comme satisfaisant si aucun des points déjà indiqués n'est caché par le masque.

Au lieu de passer en revue la façon d'obtenir mathématiquement les masques en V, nous nous contenterons d'analyser certaines règles de décision simples qui peuvent nous aider à étudier l'état de contrôle d'un processus à partir d'un graphique à somme cumulée.

En règle générale, on remarquera que le graphique à somme cumulée est essentiellement horizontal lorsque le processus est dans un état de contrôle au niveau de la valeur cible T. Si la moyenne se déplace vers le haut et prend une valeur $T_1 > T$, la somme cumulative accusera alors un mouvement ascendant ou positif. Par contre, si la moyenne se déplace vers le bas et prend une valeur $T_2 < T$, un mouvement descendant ou négatif prendra forme. Par conséquent, une tendance vers le haut ou vers le bas des points du graphique indique un changement dans la moyenne du processus, et il est alors recommandé de chercher certaine(s) cause(s) assignable(s).

Nous décrirons ci-après deux méthodes simples pour déterminer quand le graphique à somme cumulée accuse un tournant significatif.

Méthode 1:

Tracez une ligne droite à main levée qui relie les points du graphique à somme cumulée, en prolongeant cette ligne dans le sens qu'elle aurait pris s'il n'y avait pas eu de déviation manifeste vers le haut, c.-à-d., estimez la valeur moyenne des points avant le changement pressant (se reporter à la Figure 5.5). Indiquez, à cet endroit sur le graphique, un point A à une distance de 5σ au-dessus de la valeur de la somme cumulée, où σ correspond à l'écart-type de la variabilité à court terme de la série selon l'estimation donnée plus haut. On obtient ainsi **l'intervalle de décision.** Puis, indiquez un point B, dix intervalles plus loin sur le graphique, à une distance de 10σ au-dessus de l'orientation que la droite aurait prise s'il n'y avait pas eu de déviation. Tracez la ligne de décision AB. Si les points du graphique se retrouvent au-dessus de la ligne AB, c'est qu'il y a confirmation d'un changement au point pressenti. À moins que les points du graphique ne croisent cette ligne de décision avant le point de changement pressenti suivant,

le graphique n'indiquera pas de changement significatif au point A. Pour tout autre indication de changement dans la pente de la droite reliant les points du graphique, on tracera semblablement une nouvelle ligne de décision et on évaluera l'importance de ce changement pressenti en se basant sur cette ligne de référence.

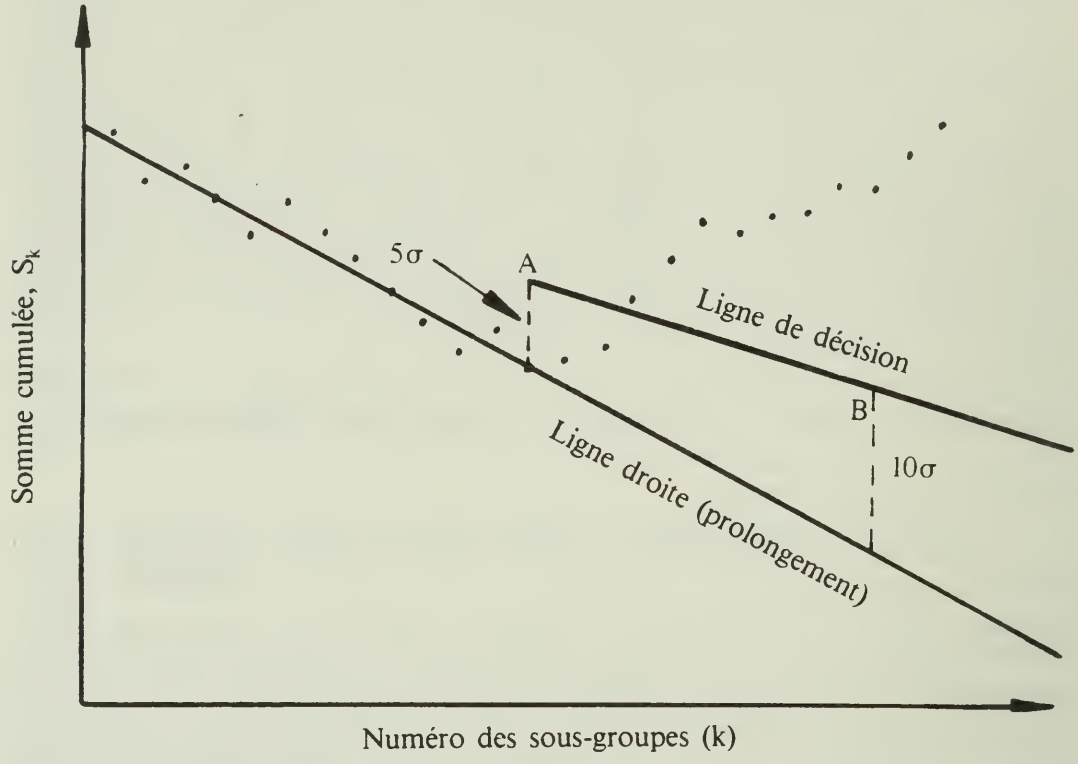


Figure 5.5 Intervalle de décision: Méthode 1

Méthode 2:

Cette méthode est un prolongement de la méthode 1 et requiert la fabrication d'un masque en V tronqué. À cette fin, on peut se servir de carton ou de papier, ou encore d'une matière transparente qui pourrait être plus pratique. Les éléments qui entrent en jeu dans la construction du masque sont illustrés à la Figure 5.6. On se sert du masque en plaçant le point de repère A sur n'importe quel point du graphique; il s'agira dans la plupart des cas du dernier point indiqué sur le graphique ou du dernier point d'un segment présentant un intérêt particulier. L'axe AF est situé parallèle à l'axe des numéros de sous-groupes, les intervalles entre les sous-groupes étant d'égale longueur sur les deux axes. Si les valeurs de sommes cumulatives dépassent les limites représentées par les axes inclinés BD et CE, appelés **lignes de décision,** cela indique un écart significatif par rapport à la valeur cible. Par contre, si toutes les valeurs sont contenues à l'intérieur de ces axes, aucun changement important n'est à signaler.

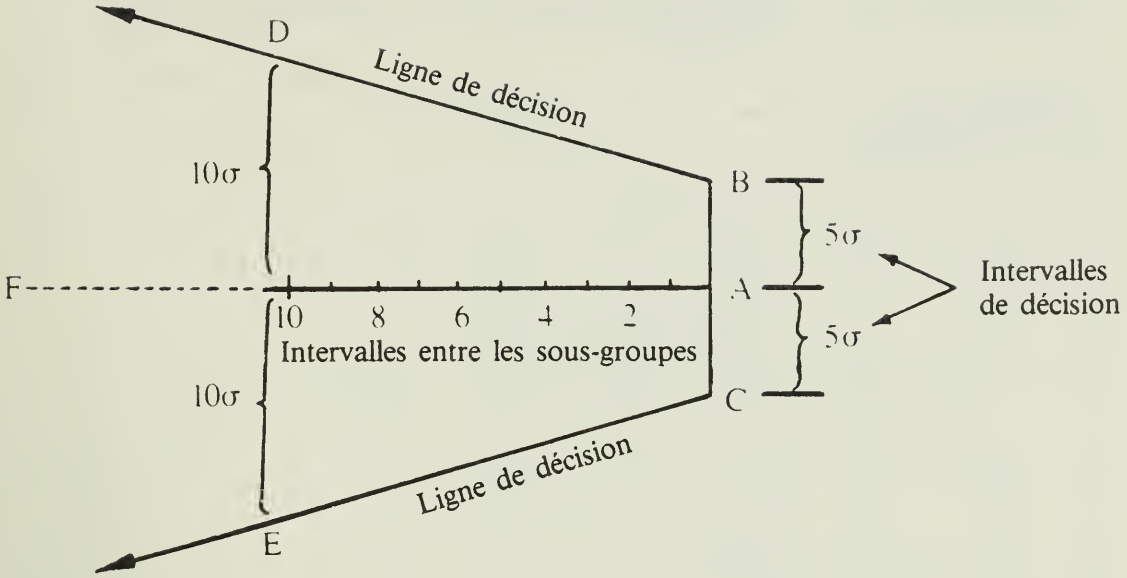


Figure 5.6 Règle de décision avec masque tronqué: Méthode 2

Exemple 5.3 Graphique à somme cumulée avec masque en V tronqué

Le Tableau 5.5 présente les valeurs moyennes de 33 sous-groupes tirés d'un procédé de production alimentaire qu'on a observé pendant une période donnée. On a aussi déterminé que T = 15 représentait la valeur cible.

TABLEAU 5.5: Données soumises à l'analyse des sommes cumulées

Numéro des sous-groupes (k)	$\bar{X}$	$\bar{X} - T$	$S_k = \Sigma(\bar{X} - T)$
1	12	−3	−3
2	17	+2	−1
3	14	−1	−2
4	14	−1	−3
5	17	+2	−1
6	16	+1	0
7	14	−1	−1
8	11	−4	−5
9	13	−2	−7
10	14	−1	−8
11	15	0	−8
12	11	−4	−12
13	14	−1	−13
14	16	+1	−12
15	13	−2	−14
16	14	−1	−15
17	11	−4	−19
18	12	−3	−22
19	13	−2	−24
20	16	+1	−23
21	12	−3	−26
22	18	+3	−23
23	18	+3	−20
24	17	+2	−18
25	20	+5	−13
26	15	0	−13
27	14	−1	−14
28	18	+3	−11
29	20	+5	−6
30	16	+1	−5
31	18	+3	−2
32	14	−1	−3
33	16	+1	−2

Le graphique à somme cumulée est illustré à la Figure 5.7. Afin de pouvoir faire une analyse comparative, on trouvera aussi à la Figure 5.8 un graphique de Shewhart type.

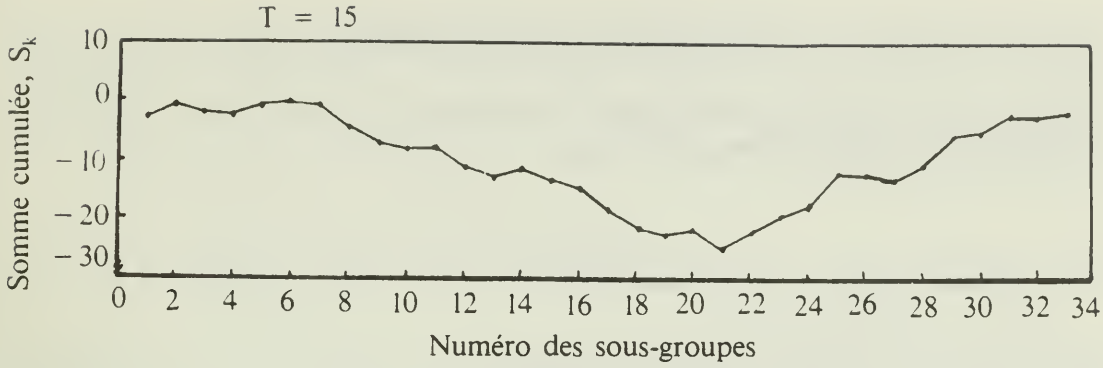


Figure 5.7 Graphique à somme cumulée des données du Tableau 5.5

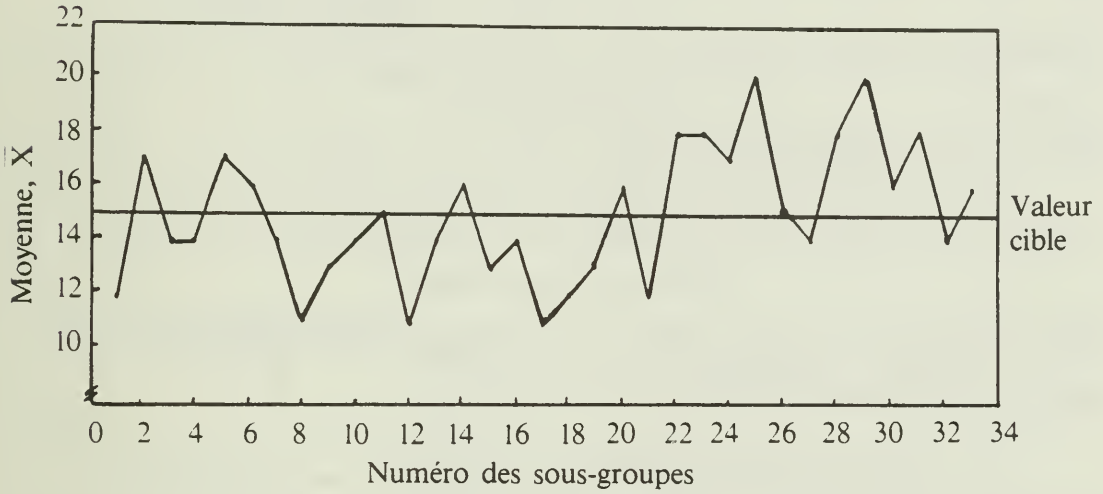


Figure 5.8 Graphique type des données du Tableau 5.5

Comme on peut le constater sur la Figure 5.7, le graphique est divisé en trois segments distincts. Du sous-groupe 1 au sous-groupe 7, la courbe des sommes cumulées est presque horizontale par rapport au zéro (0) de l'axe vertical, ce qui laisse supposer que ces observations sont tirées d'une population dont la moyenne est proche de la valeur cible. Du sous-groupe 8 au sous-groupe 21, la courbe accuse une nette tendance vers le bas; les observations semblent donc provenir d'une population dont la moyenne est inférieure à 15. Dans le troisième segment, du sous-groupe 22 au sous-groupe 33, la courbe remonte indiquant que les observations appartiennent à une population dont la moyenne est supérieure à 15.

En revanche, le graphique de Shewhart type nous permet uniquement d'affirmer que les douze dernières valeurs sont groupées autour d'une moyenne différente de celle des 20 premières environ. Un graphique à somme cumulée illustre donc beaucoup plus clairement les tendances réelles des données.

Les Figures 5.9 et 5.10 renferment un masque en V tronqué superposé à deux points différents du graphique à somme cumulée. On donne à σ la valeur 2.0. En appliquant le masque sur le sous-groupe 16, on remarque que la courbe des sommes cumulées se situe bien en deçà des axes inclinés de sorte que le segment contenant les sous-groupes 8 à 16 ne diffère pas beaucoup de la valeur cible de T = 15. Par contre, si on applique le masque sur le sous-groupe 18, on constate

que la courbe des sommes cumulées touche la ligne de décision d'en haut; cela signifie, qu'avant même d'obtenir la valeur moyenne du dix-huitième sous-groupe, suffisamment de données étaient réunies pour signaler un mouvement descendant manifeste par rapport à la valeur cible.

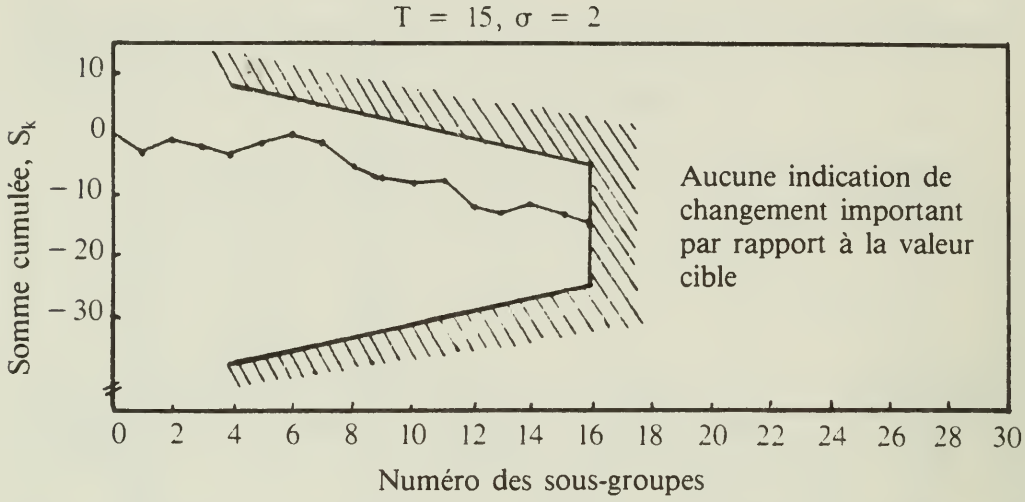


Figure 5.9 Masque en V tronqué appliqué sur le graphique à somme cumulée (Figure 5.7) au niveau du sous-groupe 16

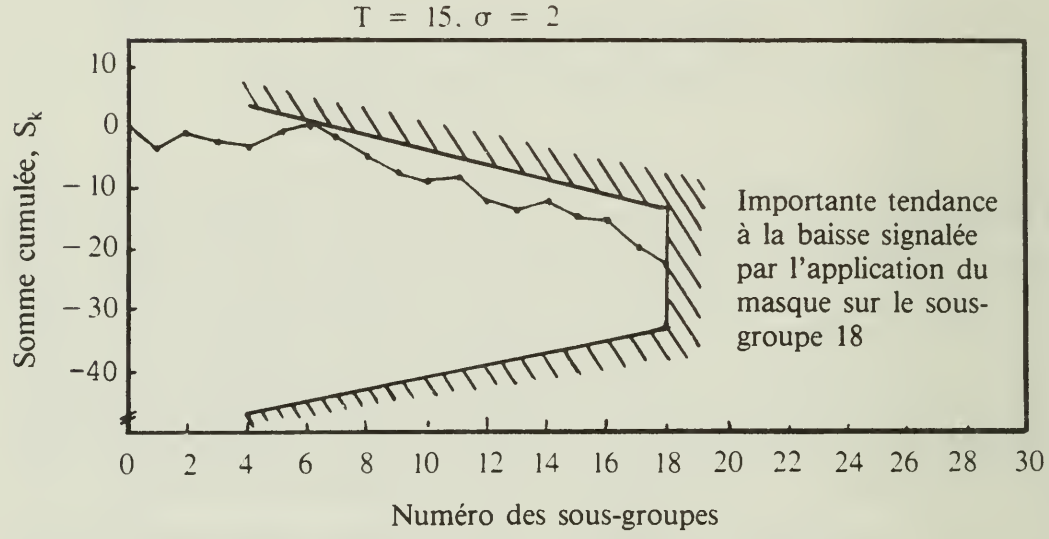


Figure 5.10 Masque en V tronqué appliqué sur le graphique à somme cumulée (Figure 5.7) au niveau du sous-groupe 18

On prendra à ce moment-ci des mesures correctives pour identifier et éliminer les causes assignables de la variabilité et, au besoin, on redémarrera le processus. La représentation graphique des données et l'analyse des sommes cumulatives se poursuivront jusqu'à ce que le processus soit dans un état de contrôle statistique.

BIBLIOGRAPHIE

1. British Standards Institution (1980): "Guide to Data Analysis and Quality Control using Cusum Techniques": BS/5703. British Standards Association, Londres, R.U.
   Part 1: Introduction to Cusum Charting
   Part 2: Decision Rules and Statistical Tests for Cusum Charts and Tabulations.

2. Association canadienne de normalisation (1985): "Programme d'assurance de la qualité": CAN3-Z299.0—Z299.4. Association canadienne de normalisation, Rexdale, Canada.

3. Craig, C.E. et Harris, R.C. (1973): "Total Productivity Measurement at the Firm Level". Sloan Management Review, p. 13-28.

4. Dodge, H.F. et Romig, H.G. (1959): "Sampling Inspection Tables, Single and Double Sampling", 2e éd.: John Wiley & Sons, New York.

5. Organisation internationale de normalisation (ISO), Secrétariat central, Genève.

   Normes internationales:

   - ISO/7002: Produits agricoles alimentaires — Plan type pour une méthode normalisée d'échantillonnage à partir d'un lot
   - ISO/8258: Graphiques de contrôle de Shewhart
   - ISO/8550: Guide pour la sélection d'un plan d'échantillonnage pour acceptation
   - ISO/8402: Qualité — Vocabulaire
   - ISO/9004: Gestion de la qualité et éléments de système qualité — Lignes directrices
   - ISO/3951: Règles et tables d'échantillonnage pour les contrôles par mesures des pourcentages de défectueux
   - ISO/3534: Statistique — Vocabulaire et symboles
     — Partie 1: Probabilité et termes statistiques généraux
     — Partie 2: Maîtrise statistique de la qualité
     — Partie 3: Plans d'expérience
   - ISO/2859: Règles d'échantillonnage pour les contrôles par attributs
     — Partie 0: Introduction générale — Plans d'échantillonnage par attributs
     — Partie 1: Plans d'échantillonnage pour les contrôles lot par lot, indexés d'après le niveau de qualité acceptable (NQA)
     — Partie 2: Plans d'échantillonnage pour les contrôles de lots isolés, indexés d'après la qualité limite (QL)
     — Partie 3: Plan d'échantillonnage successif partiel
   - ISO/8422: Plans d'échantillonnage progressif pour le contrôle par attributs

- ISO/8423: Plans d'échantillonnage progressif pour le contrôle par mesurage de la proportion d'individus non-conformes (écart-type connu)

6. Juran, J.M. (1974): "Quality Control Handbook". McGraw-Hill Book Co., New York.
7. Page, E.S. (1961): "Cumulative Sum Charts". Technometrics, Vol. 3, n° 1, p. 1-9.
8. Puri, S.C. (1984): "Statistical Process Quality Control — Key to Productivity". Standards-Quality Management Group, Ottawa, Canada.
9. Puri, S.C. (1981): "Statistical Aspects of Food Quality Assurance". Publication Numéro 5140. Agriculture Canada, Ottawa, Canada.
10. Puri, S.C. et Mullen, K. (1980): "Applied Statistics for Food and Agricultural Scientists". G.K. Hall & Co., Boston, Massachusetts.
11. Puri, S.C., Ennis, D. et Mullen, K. (1979): "Statistical Quality Control for Food and Agricultural Scientists". G.K. Hall & Co., Boston, Massachusetts.
12. Puri, S.C. (1990): "Food Safety and Quality Control: SPC with HACCP": Transactions, 44e Annual Quality Congress — American Society for Quality Control, San Francisco, Californie.
13. Puri, S.C. (1989): "Genesis of Statistical Process Control": Transactions, Update 89 Seminar. The Institute of Environmental Science and the Society of Reliability Engineers, Ottawa, Canada.
14. Puri, S.C. (1989): "Continuous Improvement: Master Check-list": Transactions, 43e Annual Quality Congress — American Society for Quality Control, Toronto, Canada.
15. Puri, S.C. (1988): "Applied SPC and the Taguchi Approach": Transactions, 19e International Symposium on Applied Technology and Automation, Monte Carlo, Monaco.
16. Puri, S.C. (1988): "International Perspectives on Quality and Standardization": Transactions, 34e Annual Quality Forum — American Society for Quality Control, Toronto Section, Toronto, Canada.
17. Puri, S.C. (1987): "Agri-Food Business: The Years Ahead": Transactions, 2e séminaire de l'Organisation européenne pour le contrôle de la qualité — Section for QC in the Food Industry, Zurich, Suisse.
18. Puri, S.C. (1987): "Developing Countries — A National Plan for Quality-Productivity": Transactions, International Conference — International Association of Science and Technology for Development, Paris, France.
19. Puri, S.C. (1987): "Management of Food Quality: Issues and Trends": Transactions, 33e Annual Quality Forum — American Society for Quality Control, Toronto Section, Toronto, Canada.

20. Puri, S.C. (1987): "Quality Challenges and Opportunities for Developing Countries": Transactions, 31e conférence annuelle — Organisation européenne pour le contrôle de la qualité, Munich, Allemagne de l'Ouest.

21. Puri, S.C. (1987): "A Plan of Excellence for a Regulatory Agency": Transactions, 41e Annual Quality Congress — American Society for Quality Control, Minneapolis, Minnesota.

22. Puri, S.C. (1986): "A Master Plan for Quality-Productivity": Transactions, 7e Latin American Quality Control Congress, Saltillo, Mexique.

23. Puri, S.C. (1985): "Quality and Deregulation": Transactions, 40e Annual Quality Congress — American Society for Quality Control, Anaheim, Californie.

24. Puri, S.C. (1984): "Quality Registration Programs": Transactions, Congrès mondial de la qualité — Organisation européenne pour le contrôle de la qualité, Brighton, R.U.

25. Puri, S.C. (1983): "Quality, Standardization and Developing Countries": Transactions, II-Asia Pacific Congress, Mexico, Mexique.

26. Puri, S.C. (1983): "Quality Indicators for Corporate Management": Transactions, 37e Annual Quality Congress — American Society for Quality Control, Boston, Massachusetts.

27. Puri, S.C. (1983): "Quality Management and Cost-Recovery": Transactions, 37e Annual Quality Congress — American Society for Quality Control, Boston, Massachusetts.

28. Schilling, E.G. (1982): "Acceptance Sampling in Quality Control". Marcel Dekker, Inc., New York.

29. Shewhart, W.A. (1931): "Economic Control of Quality of Manufactured Product". D. Van Nostrand Co., New York.

30. Stephens, K.S. (1979): "How to Perform Continuous Sampling (CSP)", Vol. 2, American Society for Quality Control, Milwaukee, Wisconsin.

31. Stephens, K.S. (1982): "How to Perform Skip-Lot and Chain Sampling", Vol. 4, American Society for Quality Control, Milwaukee, Wisconsin.

32. Wadsworth, H.M., Stephens, K.S., et Godfrey, A.B. (1986): "Modern Methods for Quality Control and Improvement". John Wiley & Sons, New York.

33. Woodward, R.H. et Goldsmith, P.L. (1967): "Cumulative Sum Techniques". Monograph n° 3. Oliver and Boyd Ltd., Londres.

APPENDICE

Tableau 1 Aire sous la courbe normale des probabilités

Tableau 2 Points de pourcentage de la distribution de t

Tableau 3 Points de pourcentage de la distribution de chi carré

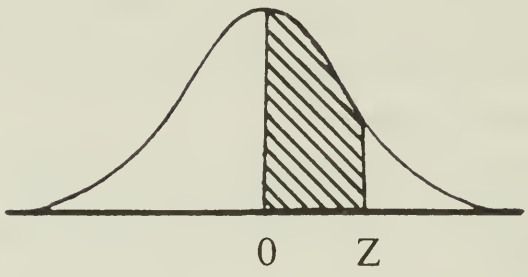
Tableau 4 Nombres aléatoires

Tableau 5A ISO/2859-1: Lettres codes des échantillons

Tableau 5B ISO/2859-1: Plans d'échantillonnage simple pour l'inspection normale

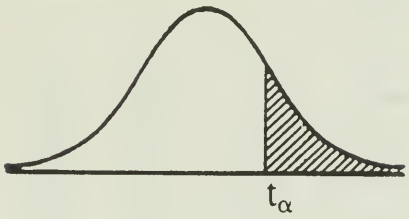
Tableau 5C ISO/2859-1: Plans d'échantillonnage double pour l'inspection normale

TABLEAU 1. Aire sous la courbe normale des probabilités



Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3883	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4992
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998

TABLEAU 2. Points de pourcentage de la distribution de t



d.f.	$t_{.40}$	$t_{.30}$	$t_{.20}$	$t_{.10}$	$t_{.05}$	$t_{.025}$	$t_{.01}$	$t_{.005}$	$t_{.0005}$
1	0.3250	0.7270	1.376	3.078	6.3138	12.706	31.821	63.657	636.619
2	.2885	.6172	1.061	1.886	2.9200	4.3027	6.965	9.9248	31.598
3	.2766	.5840	.978	1.638	2.3534	3.1825	4.541	5.8409	12.924
4	.2707	.5692	.941	1.533	2.1318	2.7764	3.747	4.6041	8.610
5	.2672	.5598	.920	1.476	2.0150	2.5706	3.365	4.0321	6.869
6	.2648	.5536	.906	1.440	1.9432	2.4469	3.143	3.7074	5.959
7	.2632	.5493	.896	1.415	1.8946	2.3646	2.998	3.4995	5.408
8	.2619	.5461	.889	1.397	1.8595	2.3060	2.896	3.3554	5.041
9	.2610	.5436	.883	1.383	1.8331	2.2622	2.821	3.2498	4.781
10	.2602	.5416	.879	1.372	1.8125	2.2281	2.764	3.1693	4.587
11	.2596	.5400	.876	1.363	1.7939	2.2010	2.718	3.1058	4.437
12	.2590	.5387	.873	1.356	1.7823	2.1788	2.681	3.0545	4.318
13	.2586	.5375	.870	1.350	1.7709	2.1604	2.650	3.0123	4.221
14	.2582	.5366	.868	1.345	1.7613	2.1448	2.624	2.9768	4.140
15	.2579	.5358	.866	1.341	1.7530	2.1315	2.602	2.9467	4.073
16	.2576	.5351	.865	1.337	1.7459	2.1199	2.583	2.9208	4.015
17	.2574	.5344	.863	1.333	1.7396	2.1098	2.567	2.8982	3.965
18	.2571	.5338	.862	1.330	1.7341	2.1009	2.552	2.8784	3.922
19	.2569	.5333	.861	1.328	1.7291	2.0930	2.539	2.8609	3.883
20	.2567	.5329	.860	1.325	1.7247	2.0860	2.528	2.8453	3.850
21	.2566	.5325	.859	1.323	1.7207	2.0796	2.518	2.8314	3.819
22	.2564	.5321	.858	1.321	1.7171	2.0739	2.508	2.8188	3.792
23	.2563	.5318	.858	1.319	1.7139	2.0687	2.500	2.9073	3.767
24	.2562	.5315	.857	1.318	1.7109	2.0639	2.492	2.7969	3.745
25	.2561	.5312	.856	1.316	1.7081	2.0595	2.485	2.7874	3.725
26	.2560	.5309	.856	1.315	1.7056	2.0555	2.479	2.7787	3.707
27	.2559	.5307	.855	1.314	1.7033	2.0518	2.473	2.7707	3.690
28	.2558	.5304	.855	1.313	1.7011	2.0484	2.467	2.7633	3.674
29	.2557	.5302	.854	1.311	1.6991	2.0452	2.462	2.7564	3.659
30	.2556	.5300	.854	1.310	1.6973	2.0423	2.457	2.7500	3.616
35	.2553	.5292	.8521	1.3062	1.6896	2.0301	2.438	2.7239	3.5919
40	.2550	.5286	.8507	1.3031	1.6839	2.0211	2.423	2.7045	3.5511
45	.2549	.5281	.8497	1.3007	1.6794	2.0141	2.412	2.6896	3.5207
50	.2547	.5278	.8489	1.2987	1.6759	2.0086	2.403	2.6778	3.4965
60	.2545	.5272	.8477	1.2959	1.6707	2.0003	2.390	2.6603	3.4606
70	.2543	.5268	.8468	1.2938	1.6669	1.9945	2.381	2.6480	3.4355
80	.2542	.5265	.8462	1.2922	1.6641	1.9901	2.374	2.6388	3.4169
90	.2541	.5263	.8457	1.2910	1.6620	1.9867	2.368	2.6316	3.4022
100	.2540	.5261	.8452	1.2901	1.6602	1.9840	2.364	2.6260	3.3909
120	.2539	.5258	.8446	1.2887	1.6577	1.9799	2.358	2.6175	3.3736
140	.2538	.5256	.8442	1.2876	1.6558	1.9771	2.353	2.6114	3.3615
160	.2538	.5255	.8439	1.2869	1.6545	1.9749	2.350	2.6070	3.3527
180	.2537	.5253	.8436	1.2863	1.6534	1.9733	2.347	2.6035	3.3456
200	.2537	.5252	.8434	1.2858	1.6525	1.9719	2.345	2.6006	3.3400
∞	.2533	.5244	.8416	1.2816	1.6449	1.9600	2.326	2.5758	3.2905

SOURCE: Tableau tiré de *Documenta Geigy Scientific Tables*, 7e édition, avec l'autorisation de la CIBA-GEIGY Limitée, Bâle, Suisse.

TABLEAU 3. Points de pourcentage de la distribution de chi carré

d.f.	$\chi^2_{.995}$	$\chi^2_{.975}$	$\chi^2_{.9}$	$\chi^2_{.5}$	$\chi^2_{.1}$	$\chi^2_{.05}$	$\chi^2_{.025}$	$\chi^2_{.01}$	$\chi^2_{.005}$	$\chi^2_{.001}$	d.f.
1	0.000	0.000	0.016	0.455	2.706	3.841	5.024	6.635	7.879	10.828	1
2	0.010	0.051	0.211	1.386	4.605	5.991	7.378	9.210	10.597	13.816	2
3	0.072	0.216	0.584	2.366	6.251	7.815	9.348	11.345	12.838	16.266	3
4	0.207	0.484	1.064	3.357	7.779	9.488	11.143	13.277	14.860	18.467	4
5	0.412	0.831	1.610	4.351	9.236	11.070	12.832	15.086	16.750	20.515	5
6	0.676	1.237	2.204	5.348	10.645	12.592	14.449	16.812	18.548	22.458	6
7	0.989	1.690	2.833	6.346	12.017	14.067	16.013	18.475	20.278	24.322	7
8	1.344	2.180	3.490	7.344	13.362	15.507	17.535	20.090	21.955	26.124	8
9	1.735	2.700	4.168	8.343	14.684	16.919	19.023	21.666	23.589	27.877	9
10	2.156	3.247	4.865	9.342	15.987	18.307	20.483	23.209	25.188	29.588	10
11	2.603	3.816	5.578	10.341	17.275	19.675	21.920	24.725	26.757	31.264	11
12	3.074	4.404	6.304	11.340	18.549	21.026	23.337	26.217	28.300	32.910	12
13	3.565	5.009	7.042	12.340	19.812	22.362	24.736	27.688	29.819	34.528	13
14	4.075	5.629	7.790	13.339	21.064	23.685	26.119	29.141	31.319	36.123	14
15	4.601	6.262	8.547	14.339	22.307	24.996	27.488	30.578	32.801	37.697	15
16	5.142	6.908	9.312	15.338	23.542	26.296	28.845	32.000	34.267	39.252	16
17	5.697	7.564	10.085	16.338	24.769	27.587	30.191	33.409	35.718	40.790	17
18	6.265	8.231	10.865	17.338	25.989	28.869	31.526	34.805	37.156	42.312	18
19	6.844	8.907	11.651	18.338	27.204	30.144	32.852	36.191	38.582	43.820	19
20	7.434	9.591	12.443	19.337	28.412	31.410	34.170	37.566	39.997	45.315	20
21	8.034	10.283	13.240	20.337	29.615	32.670	35.479	38.932	41.401	46.797	21
22	8.643	10.982	14.042	21.337	30.813	33.924	36.781	40.289	42.796	48.268	22
23	9.260	11.688	14.848	22.337	32.007	35.172	38.076	41.638	44.181	49.728	23
24	9.886	12.401	15.659	23.337	33.196	36.415	39.364	42.980	45.558	51.179	24
25	10.520	13.120	16.473	24.337	34.382	37.652	40.646	44.314	46.928	52.620	25
26	11.160	13.844	17.292	25.336	35.563	38.885	41.923	45.642	48.290	54.052	26
27	11.808	14.573	18.114	26.336	36.741	40.113	43.194	46.963	49.645	55.476	27
28	12.461	15.308	18.939	27.336	37.916	41.337	44.461	48.278	50.993	56.892	28
29	13.121	16.047	19.768	28.336	39.088	42.557	45.722	49.588	52.336	58.301	29
30	13.787	16.791	20.599	29.336	40.256	43.773	46.979	50.892	53.672	59.703	30
31	14.458	17.539	21.434	30.336	41.422	44.985	48.232	52.191	55.003	61.098	31
32	15.134	18.291	22.271	31.336	42.585	46.194	49.480	53.486	56.329	62.487	32
33	15.815	19.047	23.110	32.336	43.745	47.400	50.725	54.776	57.649	63.870	33
34	16.501	19.806	23.952	33.336	44.903	48.602	51.966	56.061	58.964	65.247	34
35	17.192	20.569	24.797	34.336	46.059	49.802	53.203	57.342	60.275	66.619	35
36	17.887	21.336	25.643	35.336	47.212	50.998	54.437	58.619	61.582	67.985	36
37	18.586	22.106	26.492	36.335	48.363	52.192	55.668	59.892	62.884	69.346	37
38	19.289	22.878	27.343	37.335	49.513	53.384	56.896	61.162	64.182	70.703	38
39	19.996	23.654	28.196	38.335	50.660	54.572	58.120	62.428	65.476	72.055	39
40	20.707	24.433	29.051	39.335	51.805	55.758	59.342	63.691	66.766	73.402	40
41	21.421	25.215	29.907	40.335	52.949	56.942	60.561	64.950	68.053	74.745	41
42	22.138	25.999	30.765	41.335	54.090	58.124	61.777	66.206	69.336	76.084	42
43	22.859	26.785	31.625	42.335	55.230	59.304	62.990	67.459	70.616	77.419	43
44	23.584	27.575	32.487	43.335	56.369	60.481	64.202	68.710	71.893	78.750	44
45	24.311	28.366	33.350	44.335	57.505	61.656	65.410	69.957	73.166	80.077	45
46	25.042	29.160	34.215	45.335	58.641	62.830	66.617	71.201	74.437	81.400	46
47	25.775	29.956	35.081	46.335	59.774	64.001	67.821	72.443	75.704	82.720	47
48	26.511	30.755	35.949	47.335	60.907	65.171	69.023	73.683	76.969	84.037	48
49	27.249	31.555	36.818	48.335	62.038	66.339	70.222	74.919	78.231	85.351	49
50	27.991	32.357	37.689	49.335	63.167	67.505	71.420	76.154	79.490	86.661	50

SOURCE: Résumé avec autorisation; tiré de *Biometrika Tables for Statisticians*, Vol. 1, édité par E.S. Pearson et H.O. Hartley, Cambridge University Press (1966).

TABLEAU 4. Nombres aléatoires

93108	77033	68325	10160	38667	62441	87023	94372	06164	30700
28271	08589	83279	48838	60935	70541	53814	95588	05832	80235
21841	35545	11148	34775	17308	88034	97765	35959	52843	44895
22025	79554	19698	25255	50283	94037	57463	92925	12042	91414
09210	20779	02994	02258	86978	85092	54052	18354	20914	28460
90552	71129	03621	20517	16908	06668	29916	51537	93658	29525
01130	06995	20258	10351	99248	51660	38861	49668	74742	47181
22604	56719	21784	68788	38358	59827	19270	99287	81193	43366
06690	01800	34272	65497	94891	14537	91358	21587	95765	72605
59809	69982	71809	64984	48709	43991	24987	69246	86400	29559
56475	02726	58511	95405	70293	84971	06676	44075	32338	31980
02730	34870	83209	03138	07715	31557	55242	61308	26507	06186
74482	33990	13509	92588	10462	76546	46097	01825	20153	36271
19793	22487	94238	81054	95488	23617	15539	94335	73822	93481
19020	27856	60526	24144	98021	60564	46373	86928	52135	74919
69565	60635	65709	77887	42766	86698	14004	94577	27936	47220
69274	23208	61035	84263	15034	28717	76146	22021	23779	98562
83658	14204	09445	41081	49630	34215	89806	40930	97194	21747
78612	51102	66826	40430	54072	62164	68977	95583	11765	81072
14980	74158	78216	38985	60838	82836	42777	85321	90463	11813
63172	28010	29405	91554	75195	51183	65805	87525	35952	83204
71167	37984	52737	06869	38122	95322	41356	19391	96787	64410
78530	56410	19195	34434	83712	50397	80920	15464	81350	18673
98324	03774	07573	67864	06497	20758	83454	22756	83959	96347
55793	30055	08373	32652	02654	75980	02095	87545	88815	80086
05674	34471	61967	91266	38814	44728	32455	17057	08339	93997
15643	22245	07592	22078	73628	60902	41561	54608	41023	98345
66750	19609	70358	03622	64898	82220	69304	46235	97332	64539
42320	74314	50222	82339	51564	42885	50482	98501	02245	88990
73752	73818	15470	04914	24936	65514	56633	72030	30856	85183
97546	02188	46373	21486	28221	08155	23486	66134	88799	49496
32569	52162	38444	42004	78011	16909	94194	79732	47114	23919
36048	93973	82596	28739	86985	58144	65007	08786	14826	04896
40455	36702	38965	56042	80023	28169	04174	65533	52718	55255
33597	47071	55618	51796	71027	46690	08002	45066	02870	60012
22828	96380	35883	15910	17211	42358	14056	55438	98148	35384
00631	95925	19324	31497	88118	06283	84596	72091	53987	01477
75722	36478	07634	63114	27164	15467	03983	09141	60562	65725
80577	01771	61510	17099	28731	41426	18853	41523	14914	76661
10524	20900	65463	83680	05005	11611	64426	59065	06758	02892
93815	69446	75253	51915	97839	75427	90685	60352	96288	34248
81867	97119	93446	20862	46591	97677	42704	13718	44975	67145
64649	07689	16711	12169	15238	74106	60655	56289	74166	78561
55768	09210	52439	33355	57884	36791	00853	49969	74814	09270
38080	49460	48137	61589	42742	92035	21766	19435	92579	27683
22360	16332	05343	34613	24013	98831	17157	44089	07366	66196
40521	09057	00239	51284	71556	22605	41293	54854	39736	05113
19292	69862	59951	49644	53486	28244	20714	56030	39292	45166
79504	40078	06838	05509	68581	39400	85615	52314	83202	40313
64138	27983	84048	42631	58658	62243	82572	45211	37060	15017

SOURCE: Résumé avec autorisation; tiré de *A Million Random Digits with 100,000 Normal Deviates*, The Rand Corporation, Santa Maria, Californie.

TABLEAU 5A. ISO/2859-1: Lettres codes des échantillons

Taille du lot	Niveaux spéciaux de contrôle				Niveaux généraux de contrôle		
	S-1	S-2	S-3	S-4	I	II	III
2 à 8	A	A	A	A	A	A	B
9 à 15	A	A	A	A	A	B	C
16 à 25	A	A	B	B	B	C	D
26 à 50	A	B	B	C	C	D	E
51 à 90	B	B	C	C	C	E	F
91 à 150	B	B	C	D	D	F	G
151 à 280	B	C	D	E	E	G	H
281 à 500	B	C	D	E	F	H	J
501 à 1 200	C	C	E	F	G	J	K
1 201 à 3 200	C	D	E	G	H	K	L
3 201 à 10 000	C	D	F	G	J	L	M
10 001 à 35 000	C	D	F	H	K	M	N
35 001 à 150 000	D	E	G	J	L	N	P
150 001 à 500 000	D	E	G	J	M	P	Q
500 001 et au-dessus	D	E	H	K	N	Q	R

SOURCE: *ISO/2859-1*, avec l'autorisation de l'Organisation internationale de normalisation (ISO), Secrétariat central, Genève.

TABLEAU 5B. ISO/2859-1: Plans d'échantillonnage simple pour l'inspection normale

| Lettre code de l'échantillon | Taille de l'échantillon | Niveaux de qualité acceptable (inspection normale) | | |

TABLEAU 5C. ISO/2859-1: Plans d'échantillonnage double pour l'inspection normale

| Lettre code de l'échantillon | Échantillon | Taille de l'échantillon | Taille cumulative de l'échantillon | Niveaux de qualité acceptable (inspection normale) | | |

LIBRARY/BIBLIOTHEQUE
AGRICULTURE CANADA OTTAWA K1A 0C5
3 9073 00072447 8

DATE DUE

AUG 30 2007

GAYLORD PRINTED IN U.S.A.

Canada